FLÁVIO GIRALDELI BIANCA

AVALIAÇÃO DE TÉCNICAS ESPECTRAIS APLICADAS À REMOÇÃO DE RUÍDO EM SINAIS DE ÁUDIO MUSICAL

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Mestre em Engenharia Elétrica.
Orientador: Prof. Dr. Evandro Ottoni Teatini Salles.
Co-orientador: Prof. Dr. Klaus Fabian Côco.

VITÓRIA
2009

Dados Internacionais de Catalogação-na-publicação (CIP)
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

Bianca, Flávio Giraldeli, 1983-

B577a Avaliação de técnicas espectrais aplicadas à remoção de ruído em sinais de áudio musical / Flávio Giraldeli Bianca. – 2009.

87 f. : il.

Orientador: Evandro Ottoni Teatini Salles.

Co-Orientador: Klaus Fabian Côco.

Dissertação (mestrado) – Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Análise espectral. 2. Wavelets (Matemática). 3. Processamento de sinais - Técnicas digitais. 4. Som - Registro e reprodução - Técnicas digitais. 5. Controle de ruído. 6. Ruído musical. 7. Subtração Espectral. I. Salles, Evandro Ottoni Teatini. II. Côco, Klaus Fabian. III. Universidade Federal do Espírito Santo. Centro Tecnológico. IV. Título.

CDU: 621.3

FLÁVIO GIRALDELI BIANCA

AVALIAÇÃO DE TÉCNICAS ESPECTRAIS APLICADAS À REMOÇÃO DE RUÍDO EM SINAIS DE ÁUDIO MUSICAL

Dissertação submetida ao programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para a obtenção do Grau de Mestre em Engenharia Elétrica.

COMISSÃO EXAMINADORA

Prof. Dr. Evandro Ottoni Teatini Salles - Orientador
Universidade Federal do Espírito Santo

Prof. Dr. Klaus Fabian Côco - Co-orientador
Universidade Federal do Espírito Santo

Prof. Dr. Mário Sarcinelli Filho
Universidade Federal do Espírito Santo

Prof. Dr. Vicente Idalberto Becerra Sablón
Universidade Salesiana

“Seja você quem for, seja qual for a posição social que você tenha na vida, a mais alta ou a mais baixa, tenha sempre como meta muita força, muita determinação e sempre faça tudo com muito amor e com muita fé em Deus, que um dia você chega lá. De alguma maneira você chega lá.”

*(*Ayrton Senna da Silva)

Ao meu grande amigo Bruno Pandolfi † (1985-2009).

Asseguro-te que enquanto existir céu e terra, você será lembrado.

Agradecimentos

Antes de tudo, agradeço a todos aqueles que sempre acreditaram na minha capacidade, o que inclui alguns professores, os meus amigos e minha família.

No que diz respeito às instituições, fica aqui registrado meu muito obrigado à UFES, através do PPGEE e toda a sua equipe de profissionais que, tenho certeza, mantém este programa de pós-graduação com muito trabalho e esforço. À CAPES pelo apoio financeiro concedido, sem o qual esta pesquisa teria sido muito mais difícil.

Ao meu orientador, professor Evandro, pelo apoio ao longo desses anos. E também ao meu co-orientador, professor Klaus.

Esta pesquisa contou com a participação de várias pessoas nos testes subjetivos. Fica aqui meu muito obrigado a todas elas.

Agradeço aos professores Bene e José Geraldo pelo voto de confiança que permitiu que eu mostrasse o meu trabalho como professor de uma instituição renomada como o IFES, sem falar nos conselhos indispensáveis para que um jovem como eu assumisse uma turma de ensino superior. E essa gratidão estende-se a toda a equipe pedagógica.

Resumo

Neste trabalho avaliou-se o problema remoção de ruído em sinais unidimensionais, notadamente sinais de áudio musical. Para isto, selecionou-se três técnicas com bons resultados relatados na literatura: Wavelet Thresholding, Time-Frequency Block Thresholding, ambas baseadas na teoria de multiresolução, além da clássica técnica Spectral Subtraction, muito utilizada em sinais de voz, sendo entretanto aqui modificada para sinais de áudio musical. Além da propriedade de remoção de ruído, cada técnica foi avaliada segundo sua capacidade de não introduzir artefatos no sinal filtrado, como o ruído musical, que é caracterizado por “notas musicais” inexistentes no áudio original. A fim de avaliar os resultados, um conjunto de sinais foi selecionado para compor a etapa de testes. Primeiramente, cada técnica foi testada em sinais sintéticos que não caracterizam áudio musical. Em seguida, as técnicas foram avaliadas para áudio de música clássico-erudita e música popular. Para cada caso, os resultados foram extensamente comentados. Mediu-se também a relação Sinal/Ruído (SNR) a fim de comparar o desempenho das técnicas. Finalmente, testes subjetivos foram implementados, realizando-se entrevistas com 10 voluntários, no intuito de se avaliar a qualidade subjetiva dos resultados de cada técnica. Os resultados mostraram que técnicas que deram bons resultados para sinais sintéticos não necessariamente se adequam à qualidade musical exigida pelos ouvintes. A modificação proposta na Spectral Subtraction foi suficiente para posicioná-la, em ternos de qualidade subjetiva, entre a Time-Frequency Block Thresholding (melhores resultados) e a Wavelet Thresholding (piores resultados).

Abstract

This work evaluated the problem of noise removal in one-dimensional signals, especially music signals. Three techniques with good results reported in the literature were selected: Wavelet Thresholding, Time-Frequency Block Thresholding, both based on the multiresolution theory, in addition to the classical Spectral Subtraction technique, widely used in speech signals, which was here modified to take into account musical audio signals. In addition to the capability to remove noise, each technique was also evaluated by their capability to not introduce artifacts in the filtered signal like the musical noise, which is characterized by “musical notes” that does not exist in the original audio. In order to evaluate the results, a set of signals was selected for the tests. First, each technique was tested with synthetic signals that do not characterize musical audio. Then, the techniques were evaluated for classical and popular music. For each case, the results are extensively discussed. In each case, it was measured the Signal to Noise Ratio (SNR) in order to compare the performance of the techniques. Finally, subjective tests were implemented, by interviews with 10 volunteers, in order to evaluate the subjective quality of the results of each technique. The results showed that techniques that have yielded good results for synthetic signals, not necessarily are suitable to the musical quality demanded by listeners. Also, the modification proposed in the Spectral Subtraction technique was enough to place it, in a subjective quality rank, between Time-Frequency Block Thresholding (best results) and Wavelet Thresholding (worst results).

Sumário

Lista de Figuras

Lista de Tabelas

Nomenclatura

Siglas

Símbolo	Descrição
1D	Unidimensional
CD	*Compact Disc* (Disco Compacto)
FFT	*Fast Fourier Transform* (Transformada Rápida de Fourier)
IDFT	*Inverse Discrete Fourier Transform* (Transformada de Fourier Discreta Inversa)
LTI	*Linear Time-Invariant* (Linear e Invariante no Tempo)
MAD	*Median Absolute Deviation*
MDCT	*Modified Discrete Cosine Transform* (Transformada Discreta do Cosseno Modificada)
MLN	*Multi Level Noise*
MMSE	*Minimum Mean-Square Error* (Erro Quadrático Médio Mínimo)
MP3	*MPEG Audio Layer-3*
PC	*Personal Computer*
VAD	*Voice Active Detection*
SNL	*Single Level Noise*
SNR	*Signal to Noise Ratio* (Relação Sinal/Ruído)
SS	*Spectral Subtraction* (Subtração Espectral)
STFT	*Short-Time Fourier Transform*
SURE	*Stein Unbiased Risk Estimate*
TFBT	*Time-Frequency Block Thresholding*
WT	*Wavelet Thresholding*

Capítulo 1: Introdução

Sinais de áudio podem ser contaminados com ruídos dos mais diversos tipos, desde os inerentes ao ambiente onde estão sendo gravados até os introduzidos artificialmente pelos meios de gravação/processamento do sinal. Em alguns casos, é possível evitar ao máximo a inserção de ruído no sinal. No primeiro caso, com um ambiente de gravação isolado acusticamente. É possível também fazer uso de equipamentos com altas relações sinal/ruído, no segundo caso.

No entanto, no mundo real é muito mais frequente o caso dos ambientes de gravação não controlados e o uso de equipamentos nem sempre de boa qualidade. Somam-se a isso a deterioração natural de algumas mídias de armazenamento analógico, como as fitas cassete.

Ruídos podem afetar não apenas a qualidade do áudio, mas também a sua inteligibilidade (no caso da presença de voz). É muito comum a existência de ruído especialmente em gravações antigas, sejam elas domésticas ou de gravadoras profissionais. Num contexto mais restrito, encontram-se as gravações de áudio musical onde não há a presença de voz.

Esse problema, ao longo de vários anos, tem motivado diversas pesquisas no sentido de atenuar a presença do ruído. Devido a diversos fatores, dentre eles a natureza do ruído e as características originais do áudio afetado, nenhuma técnica é perfeitamente eficaz em todos os casos. Além disso, cada técnica, além de remover o ruído, frequentemente afeta o sinal e pode inserir outros artefatos que não existiam originalmente.

1.1 Motivação

Atualmente predominam os sistemas de gravação de áudio digital. Nestes, uma vasta gama de ruídos são evitados, principalmente os inerentes à mídia de gravação, que são inexistentes. No entanto, há um acervo musical muito extenso, da época em que existiam apenas os sistemas de gravação e armazenamento analógicos. Busca-se, hoje, a digitalização desses materiais, a fim de preservá-los. O processo de digitalização, em si, é algo relativamente simples. Contudo, boa parte desses dados está contaminada com ruídos dos mais diversos tipos, como os ruídos tradicionalmente encontrados nas mídias analógicas de gravação magnética.

Assim sendo, o tratamento/restauração desses sinais é uma etapa importante, porém, complexa. Essa complexidade advém do fato de se estar tentando restaurar um sinal cuja forma original é desconhecida. As técnicas empregadas, pela própria natureza do ruído, são estocásticas ou exigem algum conhecimento estatístico do sinal. Nos últimos anos, têm aparecido propostas de técnicas baseadas na transformada *wavelet*, aproveitando-se sua capacidade de tratar mais adequadamente sinais com características multiescalares intrínsecas.

A grande variedade de técnicas de remoção de ruídos atualmente conhecidas nem sempre trazem bons resultados quando aplicados sobre sinais que estejam sujeitos a apreciação humana, como áudio, vídeo e imagens. Isto é o que costuma-se chamar de "percepção subjetiva de qualidade". Muitas das abordagens atuais tem excelentes resultados sobre sinais gerais, porém quando aplicadas a sinais de áudio, podem se mostrar inadequadas, inserindo, por exemplo, distorções desagradáveis, perceptíveis a audição humana. Isto sugere que uma avaliação criteriosa dos métodos aplicáveis é necessária.

1.2 Objetivos

O objetivo geral desse trabalho é examinar algumas das técnicas de remoção de ruídos em áudio atualmente conhecidas, ressaltando algumas características de cada uma delas. Dentre essas técnicas têm-se as abordagens promissoras denominadas *Wavelet Thresholding* e *Time-Frequency Block Thresholding*, além da *Spectral Subtraction*, que é uma abordagem simples, porém robusta, sendo ainda muito usada.

Como objetivo específico, busca-se avaliar o quanto uma técnica que é adequada a sinais sintéticos é apropriada para ser usada em sinais reais de áudio, como uma música.

1.3 Estrutura da Dissertação

Este trabalho encontra-se dividido em 5 capítulos.

O Capítulo 1 introduz o contexto da remoção de ruídos em áudio, bem como a motivação e objetivos envolvidos nesta pesquisa.

No Capítulo 2 aborda diversos assuntos relacionados ao tema, e que serão recorrentes ao longo do texto, o que inclui definições conceituais a respeito de áudio, tipos de ruídos e a abordagem clássica de remoção de ruídos por filtragem de Wiener.

Já no Capítulo 3, cada uma das técnicas que este trabalho se propõe avaliar é apresentada. Aspectos teóricos (principalmente) e alguns de caráter prático (algoritmo) são descritos.

Cabe ao Capítulo 4 apresentar os mais diversos testes envolvidos na avaliação de cada uma das técnicas: Objetivos, Subjetivos e Computacionais. Este capítulo, o mais extenso dessa dissertação, começa com breves testes de cada abordagem aplicados a sinais 1D sintéticos (não musicais). Na sequência, após algumas considerações a respeito, passa aos testes e resultados com sinais reais, de áudio musical instrumental, que é o foco deste trabalho.

Ao Capítulo 5, final, cabem as discussões gerais a respeito do desempenho das técnicas, com seus pontos positivos e negativos que puderam ser evidenciados ao longo de toda a pesquisa.

Capítulo 2: Conceitos Básicos

Neste capítulo serão descritos alguns conceitos que servirão de base ao entendimento deste trabalho, como aspectos relativos a áudio digital, sinais sintéticos e os mais diversos tipos de ruído.

2.1 Som

Para muitos de nós, o som é um fenômeno muito familiar, uma vez que ouvimos todo o tempo. De forma intuitiva, podemos definir som como uma sensação detectada pelos nossos ouvidos e interpretada pelo cérebro de alguma maneira. Do ponto de vista científico, som é uma perturbação física num meio (normalmente o ar). Ele se propaga neste meio como uma onda de pressão (longitudinal) através do movimento dos átomos ou moléculas.

Como qualquer outra onda, amplitude e frequência são dois importantes atributos do som. Define-se frequência como o número de períodos que ocorre em uma unidade de tempo (um segundo). A unidade de frequência é o Hertz (Hz). Um fato muito importante a ser observado é que o ouvido humano é sensível apenas à faixa de frequência em torno de 20 a 22.000 Hz, a depender da idade e saúde do indivíduo. Chama-se essa faixa de frequências audíveis. Quanto à amplitude, os humanos a percebem como o volume do som. Como o ouvido humano é sensível a uma faixa muito ampla de níveis sonoros (amplitudes), usa-se uma escala logarítmica por ser mais conveniente. A unidade resultante é o dB (decibel) [1].

2.2 Áudio Digital

Assim como uma imagem pode ser digitalizada e separada em unidades “atômicas” chamadas *pixels*, onde cada *pixel* é um número, o som também pode ser digitalizado e transformado em números. Quando o som é capturado por um microfone, ele é convertido em tensão elétrica que varia continuamente com o tempo. Tal tensão é uma representação *analógica* do som. A fim de poder ser facilmente manipulada por um computador digital, esta

representação precisa ser convertida para o domínio digital. A digitalização do som (também conhecida como conversão A/D – *Analog to Digital*) é feita através da medição da tensão elétrica em vários pontos no tempo, traduzindo-se cada medida num número, e esse conjunto de pontos é então escrito num arquivo. A esse processo dá-se o nome de amostragem (*sampling*).

Ao número de pontos, amostras (*samples*), tomados por unidade de tempo, damos o nome de taxa de amostragem. Segundo o Teorema de Nyquist, a taxa de amostragem deve ser no mínimo duas vezes maior que a máxima frequência presente no sinal analógico. Esse é o motivo pelo qual o áudio padrão do CD (*Compact Disc*) é amostrado a 44.100 Hz e quantizado a 16 bits, o que cobre uma faixa de frequência que se estende até pouco mais de 20.000 Hz, compatível com a faixa audível aos seres humanos [1].

2.3 Sinais Sintéticos Unidimensionais

É comum, quando se tratam de algoritmos de supressão de ruídos, o teste da técnica sobre sinais unidimensionais ditos “sintéticos” [2]. O termo sintético advém do fato de que tais sinais não são capturados do mundo real, mas gerados “artificialmente” por computador. Após serem gerados, são contaminados por algum tipo de ruído (geralmente aleatório com distribuição normal) (Figura 2.1). São, em seguida, processados pelo algoritmo que tentará remover o ruído do sinal. Uma vez que o pesquisador possui o sinal original em mãos, a eficácia da técnica pode ser bem avaliada, confrontando-se o sinal “limpo” com o original.

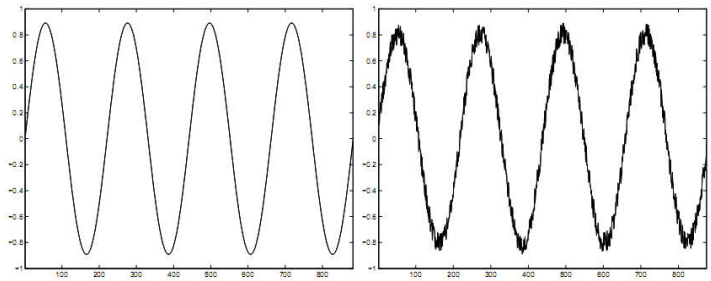

Figura 2.1 Sinal de teste original (seno de 200 Hz) (esquerda) e o mesmo contaminado com ruído gaussiano (direita).

Estes sinais são, geralmente, gerados com características importantes à avaliação da técnica, tais como continuidade, transições abruptas e/ou suaves, múltiplas frequências, etc. Dessa forma, podem-se exibir as vantagens/deficiências da técnica proposta frente às outras.

2.4 Sinais Sintéticos Unidimensionais x Áudio Digital

Apesar de se mostrarem ferramentas muito úteis na avaliação de técnicas de remoção de ruídos, uma técnica com bons resultados sobre sinais sintéticos (que não caracterizam áudio) não garantidamente terá resultados equivalentes com sinais sonoros reais, como uma música (fato este que será demonstrado neste trabalho). Não é uma tarefa fácil explicar o porquê disso, visto que tal fato está relacionado com fenômenos neurológicos e psicológicos complexos, ligados à percepção humana do som.

Em sinais de áudio, as características mais importantes consistem de oscilações [3]. Aliás, a própria "essência" dos sons advém da forma como as mais diversas oscilações ocorrem. Ruídos são também oscilações. O difícil é decidir que tipos de oscilações são ruídos e quais não. Em determinados cenários, o que matematicamente poderia ser considerado ruído é na verdade um sinal "legítimo" inserido propositalmente pelo artista. Um bom exemplo disso são as distorções presentes em guitarras em determinados estilos musicais (como o *heavy metal*). Por analogia, têm-se exemplos parecidos no campo das imagens. Um determinado fotógrafo ou cineasta pode, artificialmente ou não, "contaminar" suas imagens com tipos bem particulares de ruído como parte de sua expressão artística.

2.5 Uma Distinção Entre Voz e Música

O campo da remoção de ruídos em sinais de voz conta atualmente com uma vasta gama de técnicas eficazes na remoção dos mais diversos tipos de ruído. Pode-se dizer que boa parte do aprimoramento das técnicas foi impulsionado pelo avanço tecnológico nas comunicações, em especial da telefonia móvel. O uso de algoritmos de remoção de ruído faz-se necessário uma vez que os ambientes em que um telefone celular é usado são frequentemente ruidosos, como no trânsito.

As mesmas técnicas que são eficazes na remoção de ruído em sinais de voz podem não demonstrar a mesma eficácia quando aplicados sobre uma música. As principais razões advêm do fato desses algoritmos tirarem proveito de diversas características encontradas nos sinais de voz. Por exemplo, pelo fato do trato vocal operar de maneira relativamente lenta (comparado aos computadores), as amostras pontuais (também chamadas de *samples*) adjacentes num sinal de voz tendem a ser altamente correlacionadas [1]. E, uma vez que o ruído muitas vezes pode ser modelado como uma sequencia não correlacionada, a remoção do

ruído é facilitada. Em outras palavras, a fala acaba por gerar uma sequência "bem comportada" ou regular (Figura 2.2).

Conforme já mencionado, o ouvido humano é sensível a uma faixa de frequências que se estende, em média, até os 22 kHz. Contudo, a faixa de frequências da voz humana é muito mais restrita, geralmente estando na faixa de 500 a 2.000 Hz [1]. Isso tem pelo menos duas implicações: sinais de voz possuem períodos longos (e tendência à periodicidade) e uma simples filtragem passabaixas pode remover totalmente o ruído que está presente em frequências mais altas (supondo uma largura de banda maior que 3 kHz), uma vez que se supõe não haver nenhuma informação útil.

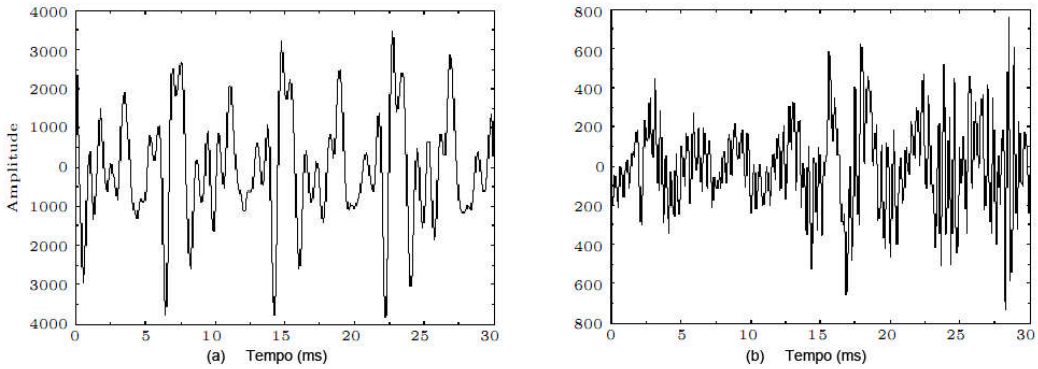

Figura 2.2 Sinais de Voz (a) e Não Voz (b) [1]

2.6 Ruído

Ruído pode ser definido como qualquer sinal indesejado que interfere na comunicação, medida, percepção ou processamento de um sinal. Ruído está presente em vários graus em quase todos os ambientes. Por exemplo, em um sistema de telefonia celular digital pode haver uma variedade de ruído que poderiam degradar a qualidade da comunicação, tais como o ruído acústico do ambiente (pessoas conversando, trânsito de veículos, etc.), ruído térmico, ruído balístico (*shot noise*), ruído eletromagnético, distorções do canal, ecos acústicos e de linha, reflexões por múltiplos caminhos e ruídos de processamento de sinal [4].

Dependendo da sua fonte, um ruído pode ser classificado em diversas categorias, indicando a ampla natureza física dos ruídos, a saber [4]:

- **Ruído Acústico** – proveniente do movimento, vibração ou colisão de fontes e é o mais familiar dentre os tipos de ruído encontrados nos ambientes do dia-a-dia. É gerado por

fontes tais como carros em movimento, ar-condicionado, tráfego, pessoas conversando ao fundo, vento e chuva;

- **Ruído Térmico** e **Ruído Balístico** (***Shot Noise***) – ruído térmico é gerado pelo movimento aleatório de partículas termicamente energizadas em um condutor elétrico. Este é intrínseco a todos os condutores e está presente mesmo sem qualquer tensão aplicada. Ruído balístico (*shot noise*) consiste de flutuações aleatórias da corrente elétrica em um condutor e é intrínseco ao fluxo de corrente. Ele é causado pelo fato da corrente ser transmitida por cargas discretas (isto é, os elétrons) com flutuações e tempos de chegada aleatórios;
- **Ruído Eletromagnético** – presente em todas as frequências e em particular na faixa de radiofrequência (kHz a GHz) onde as comunições ocorrem. Todo dispositivo elétrico, como transmissores e receptores de rádio e televisão, geram este tipo de ruído;
- **Ruído Eletrostático** – gerado pela interferência eletrostática. Lâmpadas fluorescentes são uma das fontes mais comuns de ruídos eletrostáticos;
- **Distorções de Canal**, **Eco** e **Enfraquecimento** – todos derivam das características não ideais dos canais de comunicação. Canais de rádio, tais como as frequências em GHz usadas pelas operadoras de telefonia móvel celular, são particularmente sensíveis às características de propagação do meio de transmissão;
- **Ruído de Processamento** – este é o ruído que resulta do processamento digital do sinal. Por exemplo, o ruído de quantização na codificação digital de som ou imagem, ou da perda de pacotes em sistemas de comunicação de dados, são assim classificados.

Dependendo do seu espectro de frequências ou características temporais, ruídos podem ser adicionalmente classificados em uma das seguintes categorias:

- **Ruído Branco** (***White Noise***) – ruído puramente aleatório que tem espectro de potência plano. Ruído branco contém todas as frequências em igual intensidade;
- **Ruído Branco com Largura de Banda Limitada** (***Band-Limited White Noise***) – semelhante ao ruído branco descrito acima, porém restrito a uma faixa de frequências;
- **Ruído de Banda Estreita** (***Narrowband Noise***) – um ruído com faixa de frequências estreita e limitada, tais como o "*hum*" ("zunido") de 50-60 Hz da rede elétrica;

- **Ruído Colorido** (***Coloured noise***) – ruído não branco ou qualquer outro ruído com banda larga em frequência mas espectro não plano; exemplos são o ruído rosa (Figura 2.3), marrom (Figura 2.4), e o autorregressivo;
- **Ruído Impulsivo** (***Impulsive Noise***) – consiste de pulsos de curta duração de amplitude e duração aleatória;
- **Pulsos de Ruído Transientes** (***Transient Noise Pulses***) – consistem de pulsos de ruído de duração relativamente longa. Em geral, é caracterizado por um pulso inicial curto e estreito, seguido por oscilações de mais baixa frequência (Figura 2.5).

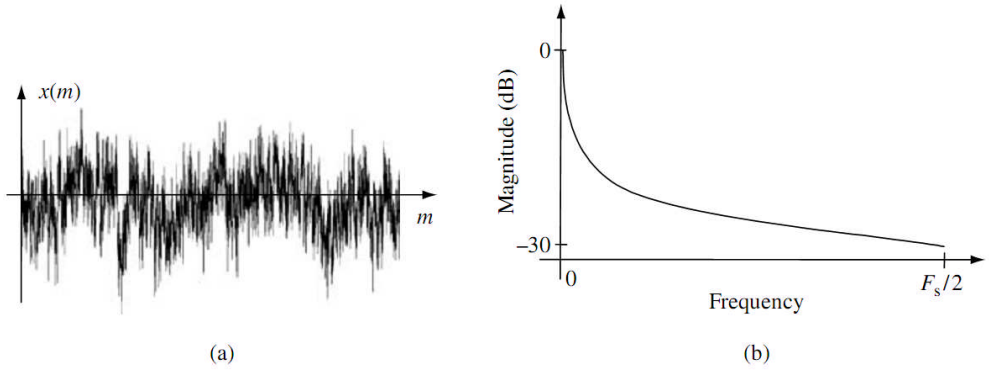

Figura 2.3 (a) Ruído Rosa (Pink Noise) e (b) seu espectro de frequências [4].

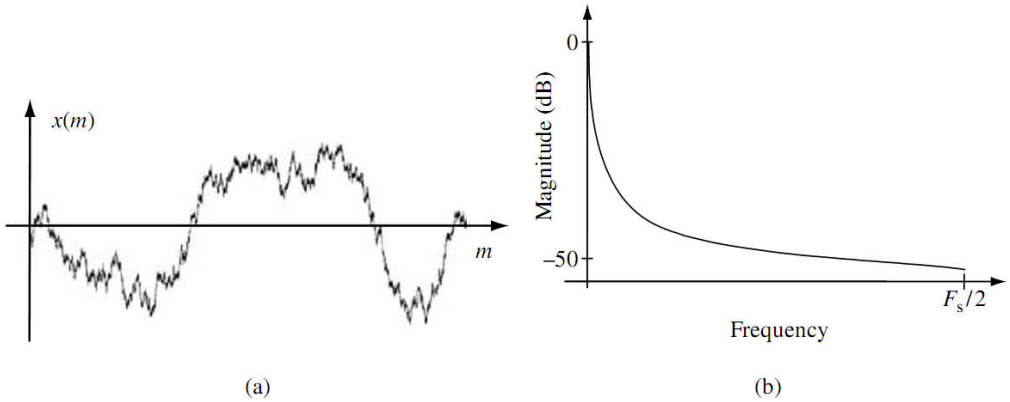

Figura 2.4 (a) Ruído Marrom (*Brown Noise*) e (b) seu espectro de frequências [4].

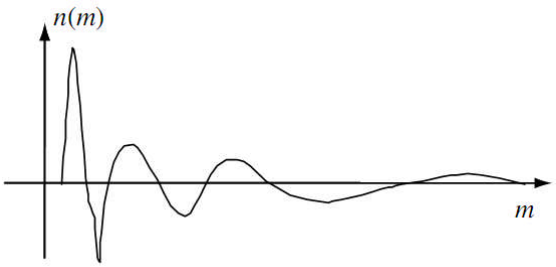

Figura 2.5 Pulsos de Ruído Transientes [4].

2.6.1 Ruído Branco

Ruído branco é definido como um processo ruidoso aleatório descorrelacionado, com potência igualmente distribuída ao longo de todas as frequências. É formado por amostras independentes e identicamente distribuídas (i.i.d.). Um ruído aleatório que tem a mesma potência em todas as frequências na faixa ±∞ precisaria necessariamente ter potência infinita; assim este é apenas um conceito teórico. Entretanto, um processo ruidoso de banda limitada, com espectro plano cobrindo a faixa de frequências de um sistema de comunicação de banda limitada, é, para todos os efeitos, do ponto de vista do sistema, um processo ruidoso branco. Por exemplo, para um sistema de áudio com largura de banda de 10 kHz, qualquer ruído sonoro de espectro plano com largura de banda igual ou superior a 10 kHz se comportará como um ruído branco [4].

A função de autocorrelação de um processo ruidoso branco de média zero $\epsilon(t)$ e no domínio do tempo contínuo com variância de σ^2 é uma função delta $\delta(\tau)$, data por

$$r_{NN}(\tau) = E[\epsilon(t)\epsilon(t+\tau)] = \sigma^2\delta(\tau) \tag{2.1}$$

O espectro de potência de um ruído branco, obtido tomando a transformada de Fourier de (2.1), é dado por

$$P_{NN}(f) = \int_{-\infty}^{\infty} r_{NN}(t)e^{-j2\pi ft}dt = \sigma^2. \tag{2.2}$$

A Equação 2.1 e a Figura 2.6(c) mostram que o ruído branco tem espectro de potência constante.

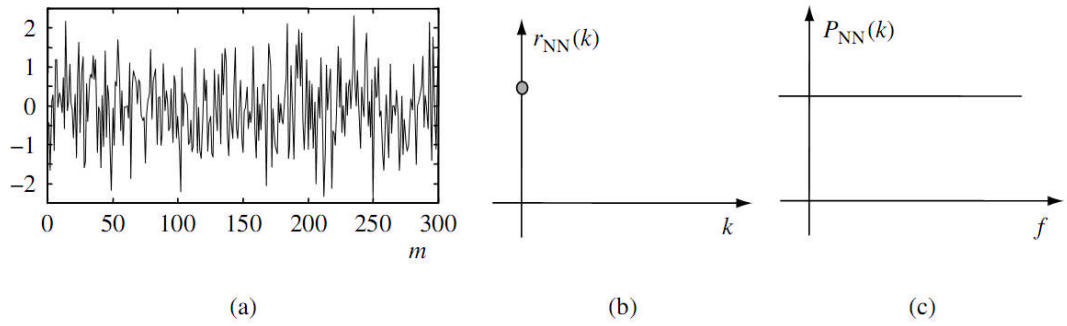

Figura 2.6 (a) Ilustração de um ruído branco. (b) Sua função de correlação é uma função delta. (c) Seu espectro de potência é constante. [4]

2.6.2 Modelando o Ruído

O objetivo da modelagem é caracterizar as estruturas e os padrões em um sinal ou ruído. E quando se tratam de ruídos em sinais de áudio, é comum assumir que o ruído é aditivo, estacionário, branco e com distribuição gaussiana [4].

Portanto, do ponto de vista matemático, podemos supor um sinal de áudio f que está contaminado por um ruído branco ϵ que é modelado por um processo estacionário gaussiano de média zero, isto é, ϵ é independente e identicamente distribuído (i.i.d) como $N(0, \sigma^2)$, e independente de f:

$$y[n] = f[n] + \epsilon[n], \quad n = 0, 1, 2, \dots, N-1. \tag{2.3}$$

Mesmo existindo uma vasta gama de problemas envolvendo ruídos não estacionários e/ou não brancos, a hipótese acima é razoável e válida na solução de muitos problemas reais. O suporte principal da validade dessa hipótese recai sobre o Teorema Central do Limite.

Em teoria de probabilidade, o Teorema Central do Limite estabelece condições nas quais a média de um número suficientemente grande de amostras de variáveis aleatórias independentes, cada uma com média e variância finitas, será aproximadamente normalmente distribuída. Uma vez que quantidades do mundo real são frequentemente uma soma balanceada de muitos eventos aleatórios não observados, este teorema provê uma explicação parcial para a prevalência da distribuição normal de probabilidade [5].

2.6.3 Removendo o Ruído

Independentemente do tipo de informação que está contaminada por ruído (como áudio, sinais sísmicos e sinais de monitoramento de processos), o campo das pesquisas envolvendo a remoção de ruído é amplo, possuindo diversas abordagens distintas, e não é exclusividade dos dias atuais. Enquanto métodos clássicos ou seus aprimoramentos ainda se mostram eficazes numa vasta gama de situações, noutras são exigidas abordagens diferentes.

2.6.3.1 Filtragem de Wiener

Quando se fala em "métodos clássicos" de remoção de ruídos, provavelmente o mais conhecido deles seja o proposto na década de 40 pelo renomado matemático americano Norbert Wiener. Partindo da indagação: Supondo que se tenha acesso ao sinal original (ou uma estimativa deste) e do ruído, qual seria o melhor filtro, capaz de separar sinal de ruído? Com uma abordagem estatística, Wiener desenvolveu o que veio a ser conhecido como Filtro de Wiener. Seu propósito é reduzir a quantidade de ruído presente em um sinal pela comparação com uma estimativa do sinal "limpo" (ou seja, o desejado). Pouco tempo depois do trabalho de Wiener, Kolmogorov publicou independentemente a versão discreta do método. Daí o fato da teoria inteira ser frequentemente denominada Filtragem de Wiener-Kolmogorov [6].

Filtros típicos são desenvolvidos para uma resposta em frequência desejada. Entretanto, o projeto do filtro de Wiener usa uma abordagem diferente. Enquanto o primeiro

assume ter conhecimento das propriedades espectrais do sinal original e do ruído, o segundo busca um filtro LTI (*Linear Time-Invariant*) no qual a saída seria tão próxima do sinal original quanto possível. Os filtros de Wiener se caracterizam por:

1. Suposição: sinal e ruído (aditivo) são processos estocásticos estacionários com características espectrais ou autocorrelação e correlação cruzadas conhecidas.
2. Requisito: o filtro deve ser fisicamente realizável, isto é, causal (apesar de ser possível encontrar uma solução não-causal).
3. Critério de desempenho: minimização do erro quadrático médio (MMSE – *Minimum Mean-Square Error*)

Filtro de Wiener do tipo FIR para sinais discretos

A fim de encontrar os coeficientes do filtro de Wiener, consideremos um sinal $w[n]$ como entrada de um filtro de Wiener de ordem N e com coeficientes $\{a_i\}, i = 0, \dots, N$. A saída do filtro é denotada por $x[n]$ a qual é dada pela expressão:

$$x[n] = \sum_{i=0}^{N} a_i y[n-i] \tag{2.4}$$

O erro residual é denotado por $e[n]$ e é definido como $e[n] = x[n] - f[n]$, como mostra a Figura 2.7.

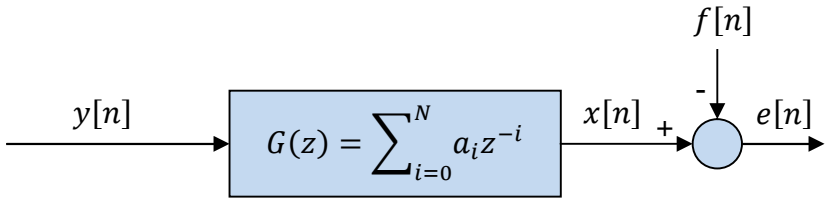

Figura 2.7 Diagrama de blocos de um filtro de Wiener do tipo FIR.

O filtro de Wiener é projetado para minimizar o erro quadrático médio (MMSE) o que pode ser estabelecido consistentemente como

$$a_i = argmin\, E\{e^2[n]\} \tag{2.5}$$

onde $E\{\cdot\}$ denota o operador de valor esperado. No caso geral, os coeficientes a_i podem ser complexos e derivados para o caso onde $y[n]$ e $f[n]$ são também complexos. Por simplicidade, consideraremos apenas o caso onde essas quantidades são reais. O erro quadrático médio pode ser escrito como

$$\begin{aligned} E\{e^2[n]\} &= E\{(x[n] - f[n])^2\} \\ &= E\{x^2[n]\} + E\{f^2[n]\} - 2E\{x[n]f[n]\} \\ &= E\{(\textstyle\sum_{i=0}^{N} a_i y[n-i])^2\} + E\{f^2[n]\} - 2E\{\textstyle\sum_{i=0}^{N} a_i y[n-i] f[n]\} \end{aligned} \tag{2.6}$$

Para encontrar o vetor $[a_0, \dots, a_N]$ que minimiza a expressão acima, vamos agora calcular sua derivada com respeito a a_i, dada por

$$\frac{\partial}{\partial a_i} E\{e^2[n]\} = 2E\left\{\left(\sum_{j=0}^{N} a_j y[n-j]\right) y[n-i]\right\} - 2E\{f[n]y[n-i]\} \quad i = 0, \dots, N$$

$$= 2\sum_{j=0}^{N} E\{y[n-j]y[n-i]\}a_j - 2E\{y[n-i]f[n]\} \tag{2.7}$$

Se nós supusermos que $y[n]$ e $f[n]$ são estacionários e conjuntamente estacionários, podemos introduzir as sequências $R_y[m]$ e $R_{yf}[m]$, conhecidas respectivamente como a autocorrelação de $y[n]$ e a correlação cruzada entre $y[n]$ e $f[n]$, definidas como

$$R_y[m] = E\{y[n]y[n+m]\} \text{ e} \tag{2.8}$$

$$R_{yf}[m] = E\{y[n]f[n+m]\}, \tag{2.9}$$

a derivada do MSE pode assim ser reescrita como (perceba que $R_{yf}[-i] = R_{fy}[i]$)

$$\frac{\partial}{\partial a_i} E\{e^2[n]\} = 2\sum_{j=0}^{N} R_y[j-i]a_j - 2R_{fy}[i], \quad i = 0, \dots, N. \tag{2.10}$$

Fazendo a derivada ser igual a zero, obtemos

$$\sum_{j=0}^{N} R_y[j-i]a_j = R_{fy}[i], \quad i = 0, \dots, N, \tag{2.11}$$

o que pode ser reescrito matricialmente como

$$\begin{bmatrix} R_y[0] & R_y[1] & \cdots & R_y[N] \\ R_y[1] & R_y[0] & \cdots & R_y[N-1] \\ \vdots & \vdots & \ddots & \vdots \\ R_y[N] & R_y[N-1] & \cdots & R_y[0] \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_N \end{bmatrix} = \begin{bmatrix} R_{fy}[0] \\ R_{fy}[1] \\ \vdots \\ R_{fy}[N] \end{bmatrix}$$

ou

$$\boldsymbol{T} \times \boldsymbol{A} = \boldsymbol{V}. \tag{2.11}$$

Essas equações são conhecidas como equações de Wiener-Hopf. A matriz $\boldsymbol{T}$ que aparece na equação é uma matriz simétrica Toeplitz. Essas matrizes são conhecidas por serem positivas definidas e, assim, não-singular, levando a uma solução única para a determinação do vetor $\boldsymbol{A}$ de coeficientes do filtro de Wiener, a saber, $\boldsymbol{A} = \boldsymbol{T}^{-1}\boldsymbol{V}$. Além disso, existe um algoritmo eficiente para resolver tais equações de Wiener-Hopt conhecido como algoritmo de Levinson-Durbin, para o qual a inversão explicita de $\boldsymbol{T}$ não é requerida.

2.6.4 Ruído Musical

Segundo Guoshen Yu [7][8], ruído musical é um artefato que é frequentemente ouvido após a aplicação de algumas técnicas de remoção de ruído, soando como notas musicais aleatórias, e tem natureza diferente (ou seja, artificial) do sinal original, podendo assim ser facilmente percebida.

Devido à falta de regularidade tempo-frequência (conceito que será melhor explicado no Capítulo 3), algoritmos ditos **diagonais** criam alguns coeficientes tempo-frequência isolados que reconstituem estruturas tempo-frequência percebidas como ruído musical (Figura 2.8). Em outras palavras, criam componentes frequenciais inexistentes no áudio original devido à excessiva quantidade de átomos tempo-frequência. Isso acontece uma vez que são usados fatores de atenuação (do ruído) não regularizados, ou seja, cada coeficiente é atenuado por um fator diferente.

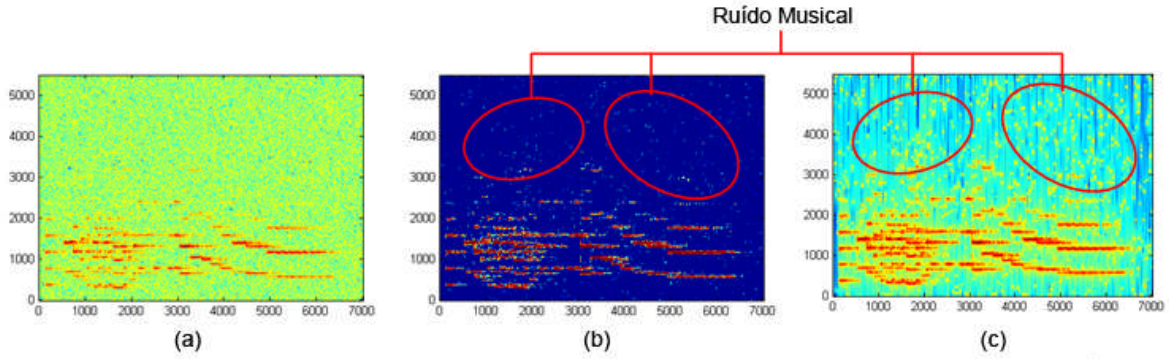

Figura 2.8 Espectrograma do sinal ruidoso (a), fatores de atenuação do ruído não regularizados (b) e o sinal filtrado exibindo os coeficientes frequenciais isolados que correspondem ao ruído musical (c) [8].

Neste sentido, diversas abordagens tentam eliminar o ruído musical, não criando esses coeficientes isolados. Isto é possível quando a técnica faz uso de estimação dita **não-diagonal**, que elimina a excessiva presença de átomos tempo-frequência neste espaço, como se estivesse regularizando-o. Assim sendo, ajustam-se tanto os coeficientes de cada átomo como também suas dimensões. O efeito é o espectro "limpo" exibido na Figura 2.9(c).

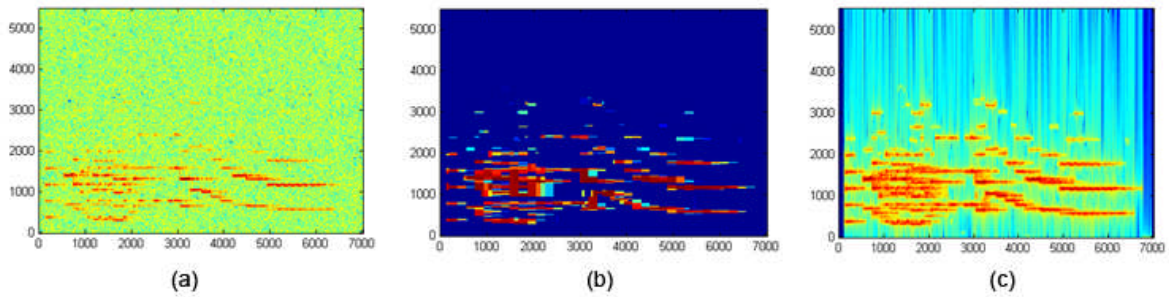

Figura 2.9 Espectrograma do sinal ruidoso (a), fatores de atenuação do ruído regularizados (b) e o sinal filtrado sem a presença do ruído musical (c) [8].

Capítulo 3: Removendo Ruído: Abordagens

Serão descritas a seguir três diferentes técnicas de remoção de ruído em áudio. Buscou-se, com essa escolha, comparar técnicas clássicas, como a *Spectral Subtraction* (década de 70) com outras modernas, como a *Wavelet Thresholding* (década de 90), bem como o teste de uma técnica que permite regularização tempo-frequência, como a *Time-Frequency Block Thresholding*.

3.1 *Spectral Subtraction*

A *Spectral Subtraction* (Subtração Espectral) é uma das mais conhecidas e populares técnicas de remoção de ruído. Artigos da década de 70 [9][10][11][12] já reportavam resultados da aplicação da técnica na remoção de ruído em sinais de voz. No entanto, a técnica original sofre de alguns problemas, que serão descritos adiante.

Em linhas gerais, suponha um sinal de áudio digital $f[n]$ corrompido por um ruído aditivo $\epsilon[n]$ [13]:

$$y[n] = f[n] + \epsilon[n], \qquad n = 0, 1, \dots, N-1. \tag{3.1}$$

Aplicando-se uma janela (janelamento[1]) e em seguida a transformada de Fourier de ambos os lados, temos

$$Y_w(e^{j\omega}) = F_w(e^{j\omega}) + \epsilon_w(e^{j\omega}), \tag{3.2}$$

onde $Y_w(e^{j\omega})$, $F_w(e^{j\omega})$ e $\epsilon_w(e^{j\omega})$ são as transformadas de Fourier de porções (janelas) do sinal ruidoso, sinal original e do ruído, respectivamente. A fim de simplificar a notação, o subscrito que indica o janelamento será retirado, ficando implícita a operação.

Multiplicando-se ambos os lados por seu complexo conjugado, tem-se

$$|Y(e^{j\omega})|^2 = |F(e^{j\omega})|^2 + |\epsilon(e^{j\omega})|^2 + 2|F(e^{j\omega})||\epsilon(e^{j\omega})|\cos(Dq), \tag{3.3}$$

onde $D(q)$ é a diferença de fase entre o sinal original e o ruído, ou seja,

$$D(q) = \angle F(e^{j\omega}) - \angle \epsilon(e^{j\omega}). \tag{3.4}$$

Tomando-se o valor esperado de ambos os lados, tem-se

$$E\left\{|Y(e^{j\omega})|^2\right\} = E\left\{|F(e^{j\omega})|^2\right\} + E\left\{|\epsilon(e^{j\omega})|^2\right\} + E\{2|F(e^{j\omega})||\epsilon(e^{j\omega})|\cos(Dq)\}$$
$$= E\left\{|F(e^{j\omega})|^2\right\} + E\left\{|\epsilon(e^{j\omega})|^2\right\} + 2\,E\{|F(e^{j\omega})|\}\,E\{|\epsilon(e^{j\omega})|\}\,E\{\cos(Dq)\}. \tag{3.5}$$

[1] O termo "janelamento" é comum na área de processamento de sinais e a partir deste ponto poderá ser usado sempre que necessário.

Analisando essa última equação, podemos assumir que:

1. a magnitude do sinal original e do ruído são independentes (com base na suposição de ruído aditivo);
2. as fases do sinal original e do ruído são também independentes uma da outra e de suas magnitudes.

Baseado na equação 3.5 observa-se que existem três fontes que contribuem como sinal contaminado: devido à potência do ruído isolada, devido ao termo cruzado sinal-ruído dado por $2\,E\{|F(e^{j\omega})|\}\,E\{|\epsilon(e^{j\omega})|\}$ e devido à diferença das fases do sinal e do ruído (Eq. 3.4). Especificamente, o modo como se trata o termo $E\{\cos(Dq)\}$ (se igual a 0 ou igual a 1) define duas variações da técnica conhecidas como Subtração Espectral de Potências, onde $E\{\cos(Dq)\} = 0$, e Subtração Espectral de Magnitudes, onde $E\{\cos(Dq)\} = 1$.

Para o caso da Subtração Espectral de Potências, o fato de $E\{\cos(Dq)\} = 0$ é consequência da hipótese de que a fase do ruído é pouco perceptível (em outras palavras, não afeta a fase do sinal final) e que, portanto, a fase do ruído pode ser desconsiderada na expressão 3.4. Então, no processamento da técnica, a fase do termo referente a $y[n]$ é assumida como a fase do termo $f[n]$, sendo desconsiderada inicialmente no processamento uma vez que apenas o espectro de potências dos sinais são manipulados. Ao final, a fase do termo relativo a $y[n]$ é adicionada ao sinal tratado e procede-se a IDFT (*Inverse Discrete Fourier Transform*) para a obtenção da estimativa do sinal no tempo.

Já na Subtração Espectral de Magnitudes é assumido que a fase do ruído pode contribuir com as distorções no sinal. Isto tem como consequência o termo $E\{\cos(Dq)\}$ tomado como 1 na expressão 3.5, mantendo, portanto, o termo cruzado $2\,E\{|F(e^{j\omega})|\}\,E\{|\epsilon(e^{j\omega})|\}$. Trabalhar com a Subtração Espectral de Magnitudes implica no conhecimento prévio (ou estimativa) da fase do ruído. Uma vez que nesta dissertação assume-se ruído branco gaussiano, optou-se por fazer $E\{\cos(Dq)\} = 0$. É digno de nota que assumir $E\{\cos(Dq)\} = 1$ torna a tarefa de remoção de ruído significativamente mais difícil [14][15][16].

Assim, como consequência de 3.5 e da hipótese de que $E\{\cos(Dq)\}$ é igual a 0, temos que

$$E\left\{|Y(e^{j\omega})|^2\right\} = E\left\{|F(e^{j\omega})|^2\right\} + E\left\{|\epsilon(e^{j\omega})|^2\right\}$$

$$|F(e^{j\omega})|^2 = |Y(e^{j\omega})|^2 - E\left\{|\epsilon(e^{j\omega})|^2\right\}. \quad (3.6)$$

Devido às flutuações presentes no espectro do ruído em torno do seu valor esperado, existirá sempre uma diferença entre o ruído real e a sua média. Além disso, poderá haver casos em que a estimativa do ruído é maior que o ruído atualmente presente no sinal. Este cenário pode levar a uma situação de picos de valores negativos no espectro resultante. O resultado desse efeito colateral, no tempo, é a percepção de um sinal artificial de múltiplas frequências, embutido no sinal filtrado. Esse ruído residual é chamado de *musical noise* ("ruído musical") na literatura [17].

Em [17] o autor propõe uma solução que minimiza esse efeito colateral. A fim de evitar possíveis valores negativos no espectro do sinal filtrado devido a uma possível "super estimativa" do espectro do ruído, o autor propõe modificações na técnica: superestimar o espectro do valor esperado do ruído antes da subtração, o que torna possível acomodar as flutuações existentes entre a estimativa e seu valor real. Ao mesmo tempo, fixar um piso (valor mínimo) para o espectro do sinal resultante da subtração. Esse valor é chamado de *spectral floor* ("piso espectral"). E, por ser expresso como uma fração do espectro do ruído, seus componentes frequenciais (coeficientes) são sempre positivos.

A fim de acomodar as variações do ruído real presente no sinal contaminado, o valor esperado do ruído (sua média), $E\left\{\left|\epsilon\left(e^{j\omega}\right)\right|^2\right\}$, é ajustado de tempos em tempos, de acordo com a estimativa da relação sinal/ruído.

Tomando $P_s(\omega)$ e $P_n(\omega)$ como o espectro em potência do sinal ruidoso e o da estimativa do ruído, e assumindo que $D(\omega) = P_s(\omega) - \alpha P_n(\omega)$, podemos enunciar a seguinte regra:

$$P_s'(\omega) = \begin{cases} D(\omega), & se\ D(\omega) > \beta P_n(\omega) \\ \beta P_n(\omega), & caso\ contrário \end{cases}$$

$$com\ \alpha \geq 1 \text{ e } 0 < \beta \ll 1$$

onde α é o fator de subtração (e controla a quantidade de ruído que será removida do sinal) e β é o parâmetro que define o piso espectral.

3.1.1 Uma Adaptação da Técnica

É importante frisar que o trabalho do autor no artigo [17] é voltado para remoção de ruídos em sinais de voz, e ainda em situações onde o ruído de fundo (*background noise*) pode mudar com o tempo. As implementações do algoritmo[2] que serviu de base a este trabalho, fazem uso de um VAD (*Voice Active Detection*) a fim de determinar em quais *frames* há a

[2] No *website* [13] recomendado pelo autor do artigo [7] há diversas implementações de algoritmos de *Spectral Subtraction* com algumas diferenças entre elas. Este trabalho baseou-se na implementação `SSBerouti79.m`.

presença de voz. O princípio de funcionamento desse algoritmo baseia-se na hipótese de que a potência do sinal de voz é sempre maior que a do ruído. Nos *frames* (trechos do sinal, que no caso da *Spectral Subtraction* contém 100 ms de áudio) onde o algoritmo sugere que não há sinal de voz, o sinal todo é frequentemente suprimido (uma vez que assume-se que é apenas ruído). E um conjunto desses *frames* é usado para atualizar a estimativa do espectro do ruído, conforme anteriormente citado.

Pode-se dizer que no contexto de filtragem de sinal de voz, as abordagens acima de fato produzem bons resultados. Contudo, na fase de pesquisa deste trabalho, diversos testes foram feitos sobre amostras extraídas de áudio musical, sem a presença de voz. E os resultados foram todos muito ruins. O sinal filtrado soava de maneira "picotada", ou seja, alternava regiões onde o som parecia estar suprimido, de outras onde a presença do ruído residual variava com o tempo. Uma análise cuidadosa dos valores intermediários calculadas pelo algoritmo, *frame* a *frame*, indicaram o porquê deste fenômeno. O fato é que no algoritmo tal qual se encontrava, quando aplicado sobre uma música onde não há presença de voz, apenas instrumentos, o VAD falha, classificando alguns *frames*, erroneamente, como apenas ruído. Esse fenômeno era especialmente perceptível em trechos onde os instrumentos tocavam de maneira sutil (baixa intensidade). Isso tinha como consequência outro efeito colateral: a estimativa do ruído tendia a divergir com o tempo (em função da atualização com os trechos classificados como apenas ruído).

Pelos motivos acima citados, a fim de ter utilidade prática no contexto desse trabalho (remoção de ruído em áudio instrumental), o algoritmo teve de ser adaptado. O primeiro passo foi remover o VAD. Mas ainda restava outro fator, o α, que controla a superestimação do ruído, ou a magnitude dos coeficientes do espectro do ruído (extraído dos *frames* iniciais), a ser analisado. No algoritmo original, esse parâmetro era automaticamente ajustado, em função das atualizações no espectro do ruído, que deixaram de existir com a desativação do VAD. Por experimentação, chegou-se a valores de α e β considerados "adequados", uma vez que balanceavam o ruído residual (ou a quantidade de ruído removido) e as distorções inseridas no sinal. Além disso, a estimativa inicial do espectro do ruído é mantida ao longo de toda a filtragem (o que é coerente com o fato de que o ruído é estacionário).

Outro parâmetro presente no algoritmo e que é típico em análise de voz é o tamanho da janela usada na análise do sinal. Esta deve ser curta o suficiente para que contenha dados de apenas um único fonema. Assim sendo, o valor comumente adotado é de 20 ms [18]. Contudo, em testes práticos com áudio musical instrumental, esse valor mostrou-se

inadequado. O fato é que janelas maiores permitem uma resolução frequencial maior[3]. Já transientes curtas precisam de janelas menores, a fim de evitar distorções. Através de diversos testes e análises de espectrogramas, adotou-se o valor de 100 ms, mais adequado ao tipo de sinal em questão.

Outra alteração feita, no campo computacional, foi um estudo das estruturadas de dados usadas, e efetuadas algumas pré-alocações que contribuíram para melhorar de maneira muito significativa (na ordem de 4x) o desempenho do algoritmo.

3.2 *Time-Frequency Block Thresholding*

Time-Frequency Block Thresholding (TFBT)[4] é uma técnica de remoção de ruídos baseada em processamento por blocos que evita a produção dos artefatos denominados "ruído musical", efeito colateral comum a muitas outras abordagens. Nesta sessão, serão descritos os aspectos principais dessa técnica proposta em [7].

TFBT busca, principalmente, regularizar a estimativa do sinal original (desconhecido). O processamento é adaptativo uma vez que os parâmetros de blocagem[5] (que ditam o tamanho dos blocos tempo-frequência) são automaticamente ajustados a partir do estimador do risco de Stein [19], calculado analiticamente a partir do sinal ruidoso.

A técnica em questão é dita ser não diagonal (conceito explicado adiante, na sessão 3.2.1.2). Esse é um conceito importante, estando envolvido com a forma como os coeficientes tempo-frequência são tratados. É preciso, portanto, estabelecer uma base para que tais conceitos possam ser melhor compreendidos.

3.2.1 Bases Teóricas da Técnica

3.2.1.1 Remoção de Ruídos em Áudio em Tempo-Frequência

Procedimentos de remoção de ruídos em áudio ditos "tempo-frequência" normalmente computam uma *Short-Time Fourier Transform* (STFT) e processam os coeficientes resultantes para atenuar o ruído. Esta transformação é executada porque revela estruturas tempo-frequência do sinal que podem ser discriminadas do ruído.

---

[3] Já que, quanto mais amostras pontuais, menor a "distância" entre as frequências representadas pelos coeficientes da Transformada de Fourier. Em outras palavras, maior a resolução frequencial.

[4] Deste ponto em diante, podendo ser referenciada apenas como *Block Thresholding*, para efeito de simplificação.

[5] Pelo termo "blocagem" entenda-se como uma denominação para o processo de agrupamento dos dados em blocos.

Suponha, para tanto, que um sinal de áudio f está contaminado com um ruído ϵ que, conforme já dito, é modelado como um processo gaussiano de média zero e independente de f, a saber

$$y[n] = f[n] + \epsilon[n], \qquad n = 0, 1, \ldots, N-1. \tag{3.7}$$

Assuma que uma transformação tempo-frequência decompõe o sinal de áudio y sobre uma família de átomos tempo-frequência $\{g_{l,k}\}_{l,k}$ onde l e k são os índices tempo e frequência, respectivamente. Os coeficientes resultantes podem então serem escritos como

$$Y[l,k] = \langle y, g_{l,k} \rangle = \sum_{n=0}^{N-1} y[n] g_{l,k}^*[n] \tag{3.8}$$

onde * denota o conjugado. Estas transformações definem uma completa e frequentemente redundante representação do sinal. Se supusermos que esses átomos tempo-frequência definem um *frame* estreito, isto significa que existe A > 0 tal que

$$\|y\|^2 = \frac{1}{A} \sum_{l,k} \left| \langle y, g_{l,k} \rangle \right|^2. \tag{3.9}$$

Isto implica numa fórmula de reconstrução simples

$$y[n] = \frac{1}{A} \sum_{l,k} Y[l,k] g_{l,k}[n]. \tag{3.10}$$

A constante A é um fator de redundância e se A = 1 então o *frame* estreito é uma base ortogonal. Um *frame* estreito comporta-se como uma união de A bases ortogonais.

Uma representação em *frame* provê um controle da energia. A redundância implica que o sinal f não tem uma forma única de ser reconstruído a partir da representação em *frame* estreito: $f[n] = (1/A) \sum_{l,k} C[l,k]\, g_{l,k}[n]$, mas todas as reconstruções satisfazem

$$\|f\|^2 \leq \frac{1}{A} \sum_{l,k} |C[l,k]|^2 \tag{3.11}$$

com uma igualdade se $C[l,k] = \langle f, g_{l,k} \rangle, \ \forall\, j,k.$

Átomos de tempo-frequência podem ser escritos como $g_{l,k}[n] = w[n - lu] exp(i2\pi kn/K)$, onde $w[n]$ é uma janela de tempo de tamanho K, a qual é deslocada com passo $u \leq K$. l e k são, respectivamente, os índices inteiros de tempo e frequência, com $0 \leq l < N/u$ e $0 \leq k < K/u$, $w[n]$ é a raiz quadrada de uma janela de *Hanning* e $u = K/2$, podendo assim ser verificado que os átomos de Fourier resultantes $\{g_{l,k}\}_{l,k}$ definem um *frame* estreito com A = 2.

Um algoritmo de remoção de ruído modifica os coeficientes tempo-frequência por multiplicar cada um deles por um fator de atenuação $a[l,k]$ para atenuar a componente de ruído. O estimador do sinal "limpo" é

$$\hat{f}[n] = \frac{1}{A} \sum_{l,k} \hat{F}[l,k] g_{l,k}[n] = \frac{1}{A} \sum_{l,k} a[l,k] Y[l,k] g_{l,k}[n]. \tag{3.12}$$

Outros algoritmos de remoção de ruídos diferem quanto ao cálculo dos fatores de atenuação $a[l,k]$. O coeficiente de variância do ruído, dado por

$$\sigma^2[l,k] = E\left\{\left|\epsilon, g_{l,k}\right|^2\right\}, \tag{3.12}$$

é supostamente conhecido, ou estimado com métodos como os referenciados em [7]. Se o ruído é estacionário, como é frequentemente o caso, então a variância do ruído não depende do tempo, ou seja, $\sigma^2[l,k] = \sigma^2[k]$.

3.2.1.2 Estimação Diagonal

Algoritmos simples de remoção de ruídos em tempo-frequência computam cada fator de atenuação $a[l,k]$ somente a partir dos componentes ruidosos $Y[l,k]$ e são por isso denominados **estimadores diagonais**. Esses algoritmos têm um desempenho limitado e produzem ruído musical. Para minimizar o limite superior do risco de estimação quadrático

$$r = E\left\{\left\|f - \hat{f}\right\|^2\right\} \leq \frac{1}{A}\sum_{l,k} E\left\{\left|F[l,k] - \hat{F}[l,k]\right|^2\right\} \tag{3.13}$$

da equação (3.13) sendo uma consequência de (3.11), pode ser demonstrado [20] que o fator de atenuação ótimo é

$$a[l,k] = 1 - \frac{1}{\xi[l,k]+1} \tag{3.14}$$

onde $\xi[l,k] = F^2[l,k]/\sigma^2[l,k]$ é o SNR (*Signal to Noise Ratio* – Relação Sinal Ruído) *a priori*. O limite inferior do risco resultante, também chamado risco do oráculo r_o, é

$$r_o \leq \frac{1}{A} R_o \tag{3.15}$$

onde

$$R_o = \sum_{l,k} \frac{|F[l,k]|^2\sigma^2[l,k]}{|F[l,k]|^2+\sigma^2[l,k]}. \tag{3.16}$$

Este limite inferior não pode ser atingido porque o fator de atenuação "oráculo", dado em (3.14) depende do SNR *a priori* $\xi[l,k]$ o qual é desconhecido. Assim, é necessário estimar este SNR.

Estimadores diagonais do SNR $\xi[l,k]$ são computados a partir do SNR *a posteriori* definido por $\gamma[l,k] = |Y[l,k]|^2/\sigma^2[l,k]$. Pode ser verificado que [7]

$$\hat{\xi}[l,k] = \gamma[l,k] - 1 \tag{3.17}$$

é um estimador não polarizado. Inserir este estimador na fórmula do oráculo (3.14) define o estimador de Wiener empírico [7]

$$a[l,k] = \left(1 - \frac{1}{\hat{\xi}[l,k]+1}\right)_+ \tag{3.18}$$

com a notação $(z)_+ = \max(z, 0)$.

O fator de atenuação $a[l,k]$ dos estimadores diagonais depende apenas de $Y[l,k]$ sem qualquer regularização tempo-frequência. Os coeficientes resultantes atenuados $a[l,k]Y[l,k]$, assim, carecem de regularidade tempo-frequência. Isto produz coeficientes tempo-frequência isolados os quais restauram estruturas tempo-frequência isoladas que são percebidas como um ruído musical. A Figura 2.8 e a Figura 2.9 ilustram esse fato.

3.2.1.2 Estimação Não Diagonal

A fim de reduzir o ruído musical bem como o risco de estimação, vários autores propuseram estimar o SNR *a priori* $\xi[l,k]$ com uma regularização tempo-frequência de um SNR *a posteriori* $\gamma[l,k]$. Os fatores de atenuação resultantes $a[l,k]$ dependerão, assim, dos valores $Y[l',k']$ para (l',k') em uma **vizinhança** inteira de (l,k), e o estimador resultante $\hat{f}[n] = (1/A)\sum_{l,k} a[l,k]Y[l,k]g_{l,k}[n]$ é dito ser não diagonal.

Estimadores não diagonais claramente superam em desempenho os estimadores diagonais, mas dependem dos parâmetros dos filtros de regularização. Filtros de regularização muito grandes reduzem a energia do ruído, mas introduzem mais distorção no sinal. Assim, é desejável que esses parâmetros de filtragem sejam ajustados dependendo da natureza do sinal de áudio. Na prática, no entanto, eles são selecionados empiricamente, com abordagem Bayesiana ou modelando o sinal como um processo Gaussiano, Gamma ou Laplaciano. Apesar de esses modelos serem frequentemente adequados para voz, eles não levam em conta a complexidade de outros sinais de áudio, como músicas, que incluem fortes ataques.

3.2.2 *Time-Frequency Block Thresholding*

Nesta sessão, a técnica, propriamente dita, será descrita. O objetivo é obter um estimador de *block thresholding* não diagonal adaptativo que automaticamente ajusta todos os parâmetros. Isto recai na habilidade de computar uma estimativa do risco, sem assumir qualquer modelo estocástico para o sinal de áudio, o que torna o algoritmo robusto.

Block Thresholding segmenta o plano tempo-frequência em blocos retangulares separados de comprimento L_i no tempo, e largura W_i na frequência. A seguir, o "tamanho do bloco" é escolhido dentre uma coleção finita de possibilidades. O *Block Thresholding* adaptativo escolhe os tamanhos por minimizar uma estimativa do risco.

O risco $r = E\left\{\left\|f - \hat{f}\right\|^2\right\}$ não pode ser calculado uma vez que f (o sinal original) é desconhecido, mas pode ser estimado com uma estimativa de risco de Stein (a ser explicado adiante). Os melhores tamanhos de blocos são computados por minimizar a estimativa do risco. O risco *block thresholding* satisfaz

$$r = E\left\{\left\|f - \hat{f}\right\|^2\right\} \leq \frac{1}{A}\sum_{i=1}^{I}\sum_{(l,k)\in B_K} E\{|a_i Y[l,k] - F[l,k]|^2\}. \quad (3.19)$$

Uma vez que $Y[l,k] = F[l,k] + \epsilon[l,k]$ e $\epsilon[l,k]$ tem média zero, $F[l,k]$ tem a mesma média de $Y[l,k]$. Para estimar o risco r, Cai [21] usa o estimador de risco de Stein quando calcula a média de um vetor aleatório, ao qual é dado pelo teorema de Stein [19] que é enunciado abaixo.

Teorema (SURE – *Stein Unbiased Risk Estimate*): Seja $\boldsymbol{Y} = \left(Y_1, \ldots, Y_p\right)$ um vetor aleatório normal com a identidade como matriz de covariância e média $\boldsymbol{F} = \left(F_1, \ldots, F_p\right)$. Seja $\boldsymbol{Y} + \boldsymbol{h}(\boldsymbol{Y})$ um estimador de $\boldsymbol{F}$, onde $\boldsymbol{h} = \left(h_1, \ldots, h_p\right) : R^p \to R^p$ quase diferenciável ($h_j : R^p \to R^1, \forall j$). Defina $\nabla \cdot \boldsymbol{h} = \sum_{j=1}^{p}\left(\partial/\partial Y_j\right)h_j$. Se $E\left\{\sum_{j=1}^{p}\left|\left(\partial/\partial Y_j\right)h_j(\boldsymbol{Y})\right|\right\} \leq \infty$, então

$$R = E\|\boldsymbol{Y} + \boldsymbol{h}(\boldsymbol{Y}) - \boldsymbol{F}\|^2 = p + E\{\|\boldsymbol{h}(\boldsymbol{Y})\|^2 + 2\nabla \cdot \boldsymbol{h}(\boldsymbol{Y})\} \quad (3.20)$$

Assim,

$$\hat{R} = p + \|\boldsymbol{h}(\boldsymbol{Y})\|_2^2 + 2\nabla \cdot \boldsymbol{h}(\boldsymbol{Y}) \quad (3.21)$$

é um estimador não polarizado do risco R de $\boldsymbol{Y} + \boldsymbol{h}(\boldsymbol{Y})$, chamado estimador de risco não polarizado de Stein.

Ou seja, com este teorema é possível reescrever qualquer expressão onde o sinal f aparece, mas é inacessível, por outra com o mesmo valor esperado, mas agora em função do sinal y contaminado por ruído (observável). É importante ressaltar, no entanto, que este resultado só se aplica aos casos de contaminação por ruído branco gaussiano.

Segundo [7] é possível particularizar o estimador do risco de Stein para cada bloco $B_i^{\#}$, de modo que a equação final[6] é dada por

$$\hat{R}_i = \bar{\sigma}_i^2\left(B_i^{\#} + \frac{\lambda^2 B_i^{\#} - 2\lambda\left(B_i^{\#} - 2\right)}{\frac{\overline{Y_i^2}}{\bar{\sigma}_i^2}}\mathbf{1}_{\overline{Y_i^2} \geq \lambda\overline{\sigma_i^2}} + B_i^{\#}\left(\frac{\overline{Y_i^2}}{\bar{\sigma}_i^2 - 2}\right)\mathbf{1}_{\overline{Y_i^2} < \lambda\overline{\sigma_i^2}}\right). \quad (3.22)$$

O *Block Thresholding* adaptativo agrupa coeficientes em blocos retangulares $B_i^{\#} = L_i \times W_i$, onde $L_i \geq 2$ e $W_i \geq 2$ são, respectivamente, o comprimento do bloco no tempo e a largura do bloco na frequência, aos quais os tamanhos são ajustados para minimizar a estimativa de risco de Stein.

A fim de regularizar a segmentação adaptativa em blocos, o plano tempo-frequência é primeiro decomposto em macroblocos M_j, $j = 1, 2, \ldots, J$, como ilustrado na Figura 3.10. Cada macrobloco M_j é segmentado em blocos B_i de mesmo tamanho, o que significa que $B_i^{\#} = P_j$ é constante sobre cada macrobloco M_j. A estimação do risco de Stein sobre M_j é $(1/$

[6] Nesta equação, o número 1 em negrito, acompanhado pela expressão condicional subscrita, indica que, caso a condição seja verdadeira, a expressão a esquerda é multiplicada por um. Caso contrário, é multiplicada por zero.

$A) \sum_{i \in M_j} \hat{R}_i$, conforme Eq. 3.21. Várias segmentações são possíveis, e nós queremos escolher aquela que leve à menor estimação do risco de Stein. O tamanho de bloco ótimo e, por conseguinte, P_j, é calculado escolhendo a forma do bloco que minimiza $\sum_{i \in M_j} \hat{R}_i$. Uma vez computados os tamanhos dos blocos, os coeficientes em cada B_i são atenuados segundo a equação (derivada a partir das equações 3.14 e 3.18)

$$a_i = \left(1 - \frac{\lambda}{\hat{\xi}_i + 1}\right)_+. \tag{3.23}$$

Assim como exposto na Eq. 3.18, o cálculo de a_i depende da estimativa do SNR *a priori* $\hat{\xi}_i$. Cada coeficiente de atenuação a_i depende de dois parâmetros, $\hat{\xi}_i$ e λ. O primeiro deles é dado pela equação

$$\hat{\xi}_i = \frac{\overline{Y_i^2}}{\overline{\sigma_i^2}} - 1, \tag{3.24}$$

onde

$$\overline{Y_i^2} = \frac{1}{B_i^{\#}} \sum_{(l,k) \in B_i} |Y[l,k]|^2, \tag{3.25}$$

$$\overline{\sigma_i^2} = \frac{1}{B_i^{\#}} \sum_{(l,k) \in B_i} \sigma^2 [l,k]. \tag{3.26}$$

Quanto ao parâmetro λ, este é calculado dependendo de $B^{\#}$ ajustando a probabilidade do ruído residual

$$Prob\{\bar{\epsilon}^2 > \lambda \sigma^2\} = \delta. \tag{3.27}$$

A probabilidade δ é um parâmetro perceptual, o qual foi escolhido como $\delta = 0.1\%$ através de experimentos psicoacústicos, valor segundo o qual o ruído musical é dificilmente perceptível (maiores detalhes podem ser encontrados em [7]).

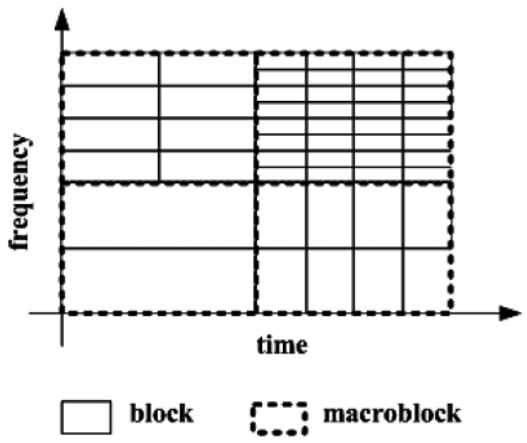

Figura 3.10 Partição de macroblocos em blocos de diferentes tamanhos [7]

Em experimentos numéricos, cada macrobloco é segmentado dentre 15 possíveis tamanhos de bloco $L \times W$, com uma combinação de $L = 8, 4, 2$ e $W = 16, 8, 4, 2, 1$. O

tamanho dos macroblocos é ajustado para ser igual ao tamanho de bloco máximo, ou seja, 8×16. A Figura 3.10 ilustra diferentes segmentações desses macroblocos em blocos tempo-frequência de mesmo tamanho. Minimizar a estimativa do risco adapta os blocos para as propriedades tempo-frequência do sinal. Em particular, ela elimina artefatos de "*pre-echo*" em sinais de ataque e resulta em menos distorções em transientes do sinal.

3.3 *Wavelet Thresholding*

De maneira análoga às análises baseadas na Transformada de Fourier, que projetam um dado sinal no domínio da frequência, os métodos que fazem uso da Transformada *Wavelet* projetam o sinal no domínio da transformada, cuja base é dada pela *wavelet* usada. Em outras palavras, no domínio da frequência (dado pela Transformada de Fourier), o sinal é representado por um somatório de funções senoidais, de diferentes fases/frequências. Já no domínio da Transformada *Wavelet*, o sinal é analogamente representado por coeficientes que quantificam a sua correlação com a função *wavelet* usada.

Donoho e Johnstone, de maneira pioneira, têm estudado a aplicabilidade das *wavelets* no campo da remoção de ruído [22]. A motivação para a aplicação da técnica advém da observação de que um sinal não aleatório, ao ser transformado, leva a um conjunto de valores esparso. Tal observação torna razoável a suposição de que alguns poucos e "grandes" coeficientes são capazes de representar bem o sinal (contêm a maior parte da informação), enquanto os demais "pequenos" coeficientes podem ser atribuídos ao ruído, que uniformemente contamina todos os coeficientes wavelet. Isso é válido para a classe de sinais contaminados por ruídos que podem ser aproximados por um conjunto de variáveis aleatórias com distribuição normal. Se pudermos decidir quais são os coeficientes *wavelets* significativos (em outras palavras, definir uma magnitude a partir da qual um coeficiente é considerado "grande"), então, os demais poderiam ser tornados iguais a zero, obtendo assim uma representação aproximada do sinal não ruidoso [22].

3.3.1 O Procedimento

Posto de maneira simplificada, a técnica básica de remoção de ruído baseada em *wavelets* é composta por três passos (cujos detalhes serão vistos posteriormente) [23]:

1. **Decomposição**: Uma vez escolhida a *wavelet* adequada à análise do sinal em questão, a decomposição é computada até um nível N determinado.

2. ***Threshold* dos coeficientes de detalhes**: Para cada um dos níveis de detalhes (1 a N), um *threshold* (fator de limiarização) é escolhido, e um procedimento de limiarização suave (*soft thresholding*) ou abrupta (*hard thresholding*) é aplicado a tais coeficientes.
3. **Reconstrução**: A transformação inversa é executada, baseada nos coeficientes modificados (ou não) do passo anterior.

Cada um desses passos guarda uma série de detalhes, muitos deles cruciais ao processo como um todo. O primeiro "problema" tem origem já na etapa de decomposição, e diz respeito à escolha da função wavelet que será usada para analisar o sinal. Um dos principais resultados que caracterizam a escolha de uma determinada wavelet como "melhor" que outra é a capacidade de gerar uma representação mais compacta do sinal, ou seja, o sinal pode ser bem aproximado com um número menor de coeficientes[7]. Isso é o mesmo que dizer que existe uma alta correlação entre a wavelet escolhida e o sinal a ser analisado. Visto que cada tipo de sinal (áudio instrumental, voz, sísmicos, etc.) guarda seu próprio conjunto de características, haverá wavelets mais ou menos apropriadas à sua análise. O problema da escolha da "melhor wavelet" traduz-se num problema mais geral, o da escolha da "melhor base", que ainda encontra-se em aberto. Em outras palavras, não existe nenhuma formulação matemática quantitativa que indique a melhor base para determinado tipo de sinal [24], em especial para áudio musical, devido a sua variabilidade.

Para o caso de análise de sinal de áudio, há alguns princípios que norteiam a escolha de uma família de wavelets mais apropriada. Visto que o requisito de fase linear é desejável [25], isto nos leva as *wavelets* biortogonais. Além disso, um alto grau de regularidade e separação das bandas de frequências é preferido (outra característica das wavelets biortogonais). Estudos com transformadas *wavelet* para áudio ou relacionadas ao tema, concordam com isso [25].

Quanto ao nível N de decomposição, deve ser escolhido de tal maneira que se consiga um bom nível de resolução frequencial, traduzido por um maior número de bandas.

É no passo 2 que a remoção de ruído acontece. Duas escolhas importantes precisam ser decididas neste momento: o tipo de *thresholding* (*soft* ou *hard*) e o valor do *threshold*. O *hard thresholding* é uma regra do tipo "mantém" ou "mata", enquanto *soft thresholding* é uma

[7] Isso é especialmente válido quando se usa análise wavelet para compressão de sinais (áudio, imagens e sinais em geral), mas é também importante no campo da remoção de ruídos, pelos motivos já citados acima.

regra do tipo "encolhe" ou "mata" (Figura 3.11). Matematicamente, poderíamos expressar essas abordagens como

$$\widehat{w}_{jk}^{*} = \delta_{\lambda}(\widehat{w}_{jk}) \tag{3.28}$$

onde tem-se que:

$\widehat{w}_{jk}$: coeficientes wavelet ruidosos

λ: valor do *threshold*

$j = 0, \dots, J-1$

$k = 0, \dots, 2^{j} - 1$

$\delta_{\lambda}(\widehat{w}_{jk}) = \widehat{w}_{jk} I(|\widehat{w}_{jk}| > \lambda)$ (*hard thresholding*),

$\delta_{\lambda}(\widehat{w}_{jk}) = sgn(\widehat{w}_{jk}) \max(0, |\widehat{w}_{jk}| - \lambda)$ (*soft thresholding*) [22].

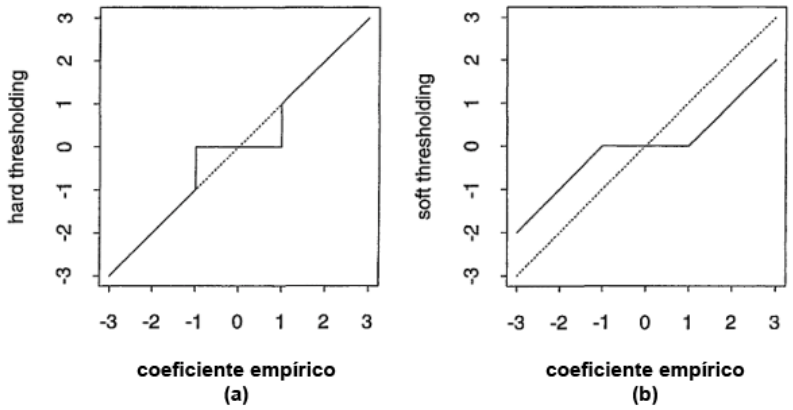

Figura 3.11 Regras (a) hard e (b) soft thresholding para λ=1 [22]

O valor do *threshold* define se um coeficiente é "zerado" ou "encolhido", nos casos *hard* ou *soft thresholding*, respectivamente. Em ambos os casos, o procedimento acontece se o valor do coeficiente está abaixo do valor decidido para o *threshold*. Podemos, assim, observar que o *thresholding* permite que os próprios dados decidam quais coeficientes wavelet são significativos. Tem sido mostrado também que *hard thresholding* resulta em uma maior variância na função de estimação, enquanto *soft thresholding* numa maior polarização [22].

Assim, a escolha do valor adequado para o *threshold* é um fator importante a ser decidido. Enquanto um valor mais alto tenderia a remover mais ruído, uma parte significativa do sinal original também poderia ser removida, e/ou, ainda, ocorrer a introdução de artefatos no sinal reconstruído[8]. Já um valor pequeno demais poderia não ter sucesso na remoção do ruído. É evidente que se está diante de um *trade off* muito presente em engenharia, onde um

[8] Conforme será comentado adiante (nas avaliações práticas de cada técnica sobre sinais de áudio "reais"), o ato de "forçar" o valor do *threshold* para um valor relativamente alto resulta em deformações no sinal reconstruído que são audivelmente mais incômodas que o próprio ruído original, cuja a remoção está sendo proposta.

valor de "meio termo" satisfatório precisa ser encontrado. Há algumas técnicas que visam encontrar o *threshold* adequado através de algumas métricas, descritas a seguir.

3.3.2 Métricas de Escolha do Valor do *Threshold*

Dentre as várias métricas existentes para a seleção do valor do *threshold*, destacam-se algumas delas a seguir. Cada uma pode ser considerada mais ou menos conservativa, dependendo da magnitude do *threshold* sugerido.

3.3.2.1 *Threshold* Universal

Donoho e Johnstone [22] propuseram o que chamaram de *threshold* universal, dado por

$$\lambda_{un} = \sigma \sqrt{2 \log(n)} / \sqrt{n}, \tag{3.29}$$

onde σ é o desvio padrão (ou uma estimativa do mesmo, caso este seja desconhecido) da massa de dados com n amostras pontuais.

Apesar da simplicidade de tal métrica, os referidos autores mostraram que tanto para *hard* quanto *soft thresholding* o estimador wavelet não-linear resultante é assintoticamente próximo ao estimador *minimax*, em termos do erro quadrático médio do *L²-risk*. Mais ainda, para funções não homogêneas ele supera qualquer estimador linear [22]. Mais informações e uma completa formulação matemática acerca desse estimador podem ser encontradas em [26].

3.3.2.2 SURE (*Stein's Unbiased Risk Estimate*)

Em [2] é proposta uma regra de seleção para o valor do *threshold* empregando-se a técnica *SureShrink*, que se baseia na minimização do "estimador de risco não polarizado de Stein" (em inglês, *Stein's Unbiased Risk Estimate*). Maiores informações podem ser obtidas em [2].

Dentre as métricas baseadas no *threshold* universal, *minimax* e SURE, esta última obteve os melhores resultados em testes preliminares de percepção subjetiva de qualidade.

3.3.3 O Procedimento: Mais Detalhes

Em implementações da técnica Wavelet Thresholding, o valor do *threshold* é, de uma maneira ou de outra, calculado a partir dos próprios dados. O tipo de informação colhida a partir dos dados, e que serve de parâmetro para o cálculo deste valor, é que varia entre uma métrica e outra.

Um detalhe ainda não mencionado a respeito do procedimento de *thresholding* dos coeficientes é que, uma vez calculado o valor do *threshold*, este pode ser aplicado diretamente aos coeficientes ou sofrer um reescalamento.

No modelo básico, nenhum reescalamento é feito. O valor de threshold calculado em cada nível é diretamente aplicado.

No modelo denominado SNL (*Single Level Noise*), o valor *threshold* sofre um reescalamento baseado numa estimação do nível do ruído, que é calculado a partir dos coeficientes de detalhe do primeiro nível de decomposição. Este "fator de reescalamento" é usado para todos os demais níveis. Segundo [23], os coeficientes de detalhe do primeiro nível são essencialmente coeficientes ruidosos com desvio padrão igual a σ. O MAD (*Median Absolute Deviation* – Desvio Absoluto Mediano) desses coeficientes é uma estimativa robusta de σ. O uso de uma estimativa robusta é crucial por duas razões: a primeira é que se os coeficientes do nível 1 contêm f detalhes, então estes detalhes estão concentrados em uns poucos coeficientes se a função é suficientemente regular; a segunda razão é para evitar "efeitos de bordas", que são puros artefatos devido às computações nas bordas.

O último dos modelos, MLN (*Multi Level Noise*), é aplicável em situações de ruído não branco, nas quais a estimativa do nível de ruído é feita em cada um dos níveis de decomposição, independentemente. A estratégia é a mesma já descrita no parágrafo anterior, porém com a diferença da estimação ser refeita nível a nível.

Na etapa de pesquisa desse trabalho, muitos testes foram feitos a fim de investigar os detalhes inerentes a essa técnica, e como cada escolha (a família de *wavelets*, o tipo de *thresholding*, de reescalamento) se reflete na qualidade do sinal processado. Uma vez escolhida a métrica de cálculo do valor do *threshold*, o valor é automaticamente obtido. No entanto, visto que em muitos testes auditivos ainda havia uma considerável quantidade de ruído residual, questionou-se o que aconteceria se o valor do *threshold* sofresse um ganho que o forçasse a ser maior. Já era de conhecimento prévio o fato de que, um valor maior removeria mais ruído, porém poderia inserir mais distorções ao sinal. O que seria mais fácil de ser tolerado: o ruído residual ou os artefatos inseridos? A resposta veio quando a técnica foi modificada a fim de forçar um ganho multiplicativo (maior que 1.0) no valor do *threshold* antes de sua aplicação aos coeficientes. E o fato é que o nível de distorções (artefatos) que o sinal sofre supera (e muito) a pouca remoção a mais de ruído proporcionada por um *threshold* maior. Isso motivou a decisão de manter o valor do *threshold* aplicado igual ao calculado pela métrica usada.

Capítulo 4: Testes Experimentais e Resultados

As influências, positivas ou negativas, de cada técnica de remoção de ruído sobre amostras[9] de teste são bastante distintas. Qualquer técnica de remoção de ruído, por mais avançada que seja, irá naturalmente inserir artefatos que não existiam no áudio original. Um aspecto interessante é que as distorções inseridas diferem não só pela técnica, mas algumas vezes pelo tipo de amostra. Em outras palavras, uma mesma técnica pode dar melhores resultados sobre um tipo de sinal que outros.

Neste capítulo, antes de proceder às análises com amostras de áudio musical, sinais sintéticos contaminados com ruídos da mesma natureza dos que serão usados nas músicas adiante, serão submetidos às técnicas de remoção de ruído. O objetivo é demonstrar como sinais reais diferem de sinais sintéticos, conforme já afirmado no Capítulo 2.

Após isso, passaremos às amostras extraídas de músicas. Inicialmente, será feita uma análise que visa caracterizar cada uma das amostras de áudio que foram, ao longo de toda a pesquisa, selecionadas para avaliar as técnicas. Tais caracterizações se darão não apenas no campo de processamento de sinais, mas envolverão alguns aspectos inerentes ao campo da música. Isso permitirá uma análise mais próxima da realidade (da percepção humana subjetiva da música).

Diferentemente de muitos textos (artigos, livros, teses, etc.) [7][17][27] envolvendo pesquisas neste mesmo campo, nesta dissertação optou-se por usar amostras longas (mais de 10 segundos) e de trechos reais de músicas. Dessa forma, espera-se que numa mesma amostra, cada técnica mostre-se mais ou menos eficiente em cada trecho.

Pelos motivos já citados em capítulos anteriores, nenhuma das amostras envolve sinal de voz, apenas áudio instrumental (porém de vários estilos musicais). Inicialmente, ao definir que a pesquisa se focaria em áudio instrumental, buscou-se extrair amostras em álbuns de música clássico-erudita, por estas serem tipicamente instrumentais[10]. Contudo, esta escolha mostrou-se não adequada, pois as estruturas que seriam cruciais na avaliação das técnicas (como transientes curtos, amplo espectro de frequências e distorções propositais) seriam

[9] Neste trabalho, os termos “amostra” ou “amostra de teste” serão frequentemente usados e ambos referem-se a trechos de músicas usados para exemplificar e testar o comportamento/desempenho de cada técnica. O termo difere, portanto, do usado no contexto de processamento de sinais, na qual é usado para denominar um único valor numérico tomado em um processo de amostragem. Para este caso, o termo usado será “amostra pontual” ou *sample*. Uma amostra ou amostra de teste seria, portanto, um conjunto de valores numéricos (amostras pontuais), e não um único valor.

[10] Salvo algumas exceções, como a conhecida 9ª Sinfonia de Beethoven, que possui um coral de vozes em seu último movimento.

difíceis de serem encontradas nesse gênero de música tão "bem comportado" (ou mais "previsível").

Uma possível prova que fundamenta essa afirmação pode ser encontrada no campo de compressão de áudio. O conhecido formato MP3, por exemplo, tende a gerar arquivos mais compactos quando se trata de músicas clássicas, ao contrário de outros gêneros populares (como pop, rock, eletrônico, etc.), supondo a configuração para o mesmo padrão de qualidade. O teste abaixo (Tabela 4.1) exibe essa diferença de maneira quantitativa [28]. O teste faz uso do *codec* LAME que, quando configurado no modo de qualidade, usa uma escala que vai de V0 (maior qualidade) a V10 (menor qualidade). Na tabela é possível observar que o *bitrate* (taxa de bits) médio de um arquivo de música clássica comprimido é sempre menor que o de outros gêneros. A explicação técnica básica e simplificada (uma vez que há diversos outros fatores mais complexos envolvidos) é que, como o princípio da compressão de áudio do MP3 é a quantização dos coeficientes resultantes de uma transformação (no caso a MDCT – *Modified Discrete Cosine Transform*) [1], pode-se afirmar que, quanto mais complexo (mais rico em componentes espectrais) o sinal, mais coeficientes significativos são gerados, e, portanto, precisam ser mantidos no processo de quantização a fim de que a qualidade seja assegurada. Mais coeficientes diferentes de zero, mais dados para serem armazenados. Assim sendo, é possível afirmar que, ao menos do ponto de vista de processamento digital de sinais, áudio musical do gênero clássico-erudito tende a ser composto por estruturas menos complexas (se comparados aos demais gêneros populares). Assim sendo, voltou-se para faixas de bandas de "música popular" contemporâneas, que se adequariam aos objetivos desejados (maior complexidade sonora).

Tabela 4.1 *Bitrate* médio (25 faixas) para músicas clássicas e outros gêneros [28]

	V10	V9.5	V9	V8.5	V8	V7.5	V7	V6.5	V6	V5.5	V5	V4.5	V4	V3.5	V3	V2.5	V2	V1.5	V1	V0.5	V0
Various	30.5	52.3	69.8	79.0	97.0	100.9	107.6	112.5	127.4	135.3	141.5	149.0	156.4	163.3	170.5	192.4	204.0	211.9	224.6	235.2	254.0
Classical	25.0	44.2	58.1	67.0	79.2	86.0	93.8	101.0	117.0	124.4	130.8	142.6	153.5	160.7	168.0	175.0	182.7	190.8	199.8	211.0	223.3
Overall av.	**27.8**	**48.2**	**64.0**	**73.0**	**88.1**	**93.5**	**100.7**	**106.7**	**122.2**	**129.8**	**136.2**	**145.8**	**154.9**	**162.0**	**169.3**	**183.7**	**193.4**	**201.4**	**212.2**	**223.1**	**238.6**

Após comentados os aspectos de cada amostra escolhida, será feita uma análise subjetiva de cada amostra tratada, pontuando as diferenças entre os resultados de cada técnica. Isso se faz importante pois não há métrica ou análise "objetiva" que consiga, por exemplo, dizer que a alguns harmônicos de um acorde se foram, e se o quanto isso influenciou a apreciação da música. Em outras palavras, não há métrica capaz de julgar esses elementos artísticos/subjetivos.

O próximo passo envolverá testes mais estatísticos, tanto subjetivos como objetivos. Quanto ao primeiro caso, algumas informações serão extraídas através da exposição de vários

ouvintes a cada uma das amostras tratadas. Já nos objetivos, serão exibidas algumas métricas tradicionalmente usadas neste campo, como a relação sinal/ruído.

Para o cálculo do SNR será adotada a seguinte fórmula, em conformidade com a implementação do autor em [7]:

$$SNR = 10\frac{log\sum_{n=1}^{N} f_o[n]^2}{log\sum_{n=1}^{N}(f_o[n]-f[n])^2} \quad (4.1)$$

onde têm-se que:

f_o: sinal original (não contaminado por ruído)

f: sinal ruidoso.

N: Número de amostras pontuais.

Além disso, caso os somatórios $\sum_{n=1}^{N} f_o[n]^2$ ou $\sum_{n=1}^{N}(f_o[n]-f[n])^2$ resultem em zero, define-se *SNR = 0*.

4.1 *Denoising* de Sinais Sintéticos

Em [2], Donoho e Johnstone, propõem seis funções de teste que mais tarde foram inclusive incorporadas ao MATLAB®, sendo referência na geração de sinais de testes para algoritmos de remoção de ruído.

A seguir, cada uma das funções de teste é contaminada com ruído gaussiano branco aditivo (variância de 0,4 e média 0)[11] e submetida a cada uma das técnicas abordadas no Capítulo 3. Os SNR's antes/depois são também calculados, que, juntamente com a inspeção visual dos sinais plotados, possibilitarão uma boa análise do desempenho de cada técnica com sinais sintéticos.

Em virtude de duas particularidades da atual implementação da *Spectral Subtraction*, o SNR é "erroneamente" calculado, pois há uma alteração (ganho) na magnitude do sinal após a aplicação da técnica e *samples* "em branco" no final do sinal, em virtude do processamento em bloco inerente à técnica. Esse último aspecto também afeta o cálculo do SNR da *Block Thresholding* em algumas amostras (aquelas que não terminam naturalmente com alguns *samples* de valor zero). Todavia, nestes casos, ainda resta a inspeção visual para aferir o desempenho da técnica.

[11] A escolha desses valores se deu de maneira experimental, de tal forma que a magnitude do ruído (em relação ao sinal original) fosse suficiente para demonstrar o desempenho de cada técnica.

4.1.1 Testes com Sinais Sintéticos

4.1.1.1 *Blocks*

A primeira função a ser testada, *blocks*, é exibida na Figura 4.12(a) bem como sua versão ruidosa em (b). Esta função gera um sinal que contém tipicamente transições rápidas representadas por degraus, e regiões onde seu valor é constante ao longo do tempo. A Tabela 4.2 confirma o excelente desempenho da *Wavelet Thresholding* na limpeza do sinal. Em virtude dos "problemas" citados no cálculo do SNR das técnicas *Block Thresholding* e *Spectral Subtraction*, o SNR acaba por não refletir o bom desempenho que essas técnicas também obtiveram, que pode ser conferido visualmente. No entanto, *Block Thresholding* lida melhor com as transições abruptas (bordas), se comparado com as demais.

Tabela 4.2 SNR antes/depois da filtragem da função *blocks*

Técnica	SNR Antes (dB)	SNR Depois (dB)
Block Thresholding	19,57	12,87
Spectral Subtraction	19,57	8,94
Wavelet Thresholding	19,57	27,84

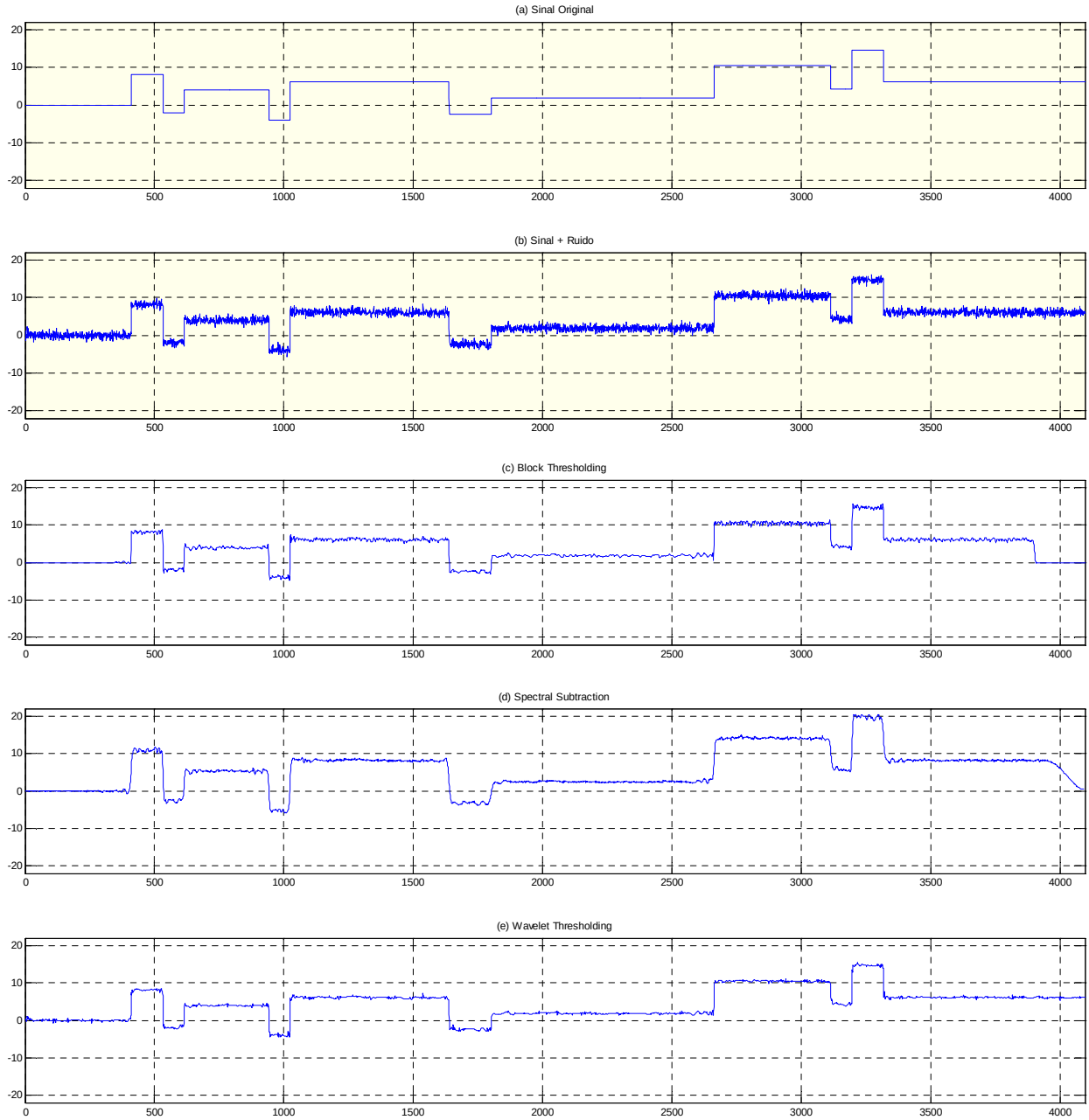

Figura 4.12 (a), (b): *Blocks* original e ruidosa. (c), (d), (e): *Blocks* tratada com cada uma das técnicas indicadas. O eixo vertical representa a magnitude e o horizontal o número de amostras pontuais.

4.1.1.2 *Bumps*

Esta é uma função de características "contrárias" à função *blocks* comentada anteriormente. *Bumps* contém exclusivamente transições suaves, porém com algumas "sutilezas", como pode ser observado na Figura 4.13(a), para $1000 \leq n \leq 1750$, onde n representa o índice da amostra pontual. Todas as técnicas executam um bom trabalho. Conforme a Tabela 4.3, *Block Thresholding* e *Wavelet Thresholding* tem SNR's quase idênticos. *Wavelet Thresholding* e *Spectral Subtraction* lidaram melhor com as sutilezas do sinal comentadas há pouco, apesar do pequeno ganho da *Spectral Subtraction* ter prejudicado a medição do seu SNR.

Tabela 4.3 SNR antes/depois da filtragem da função *bumps*

Técnica	SNR Antes (dB)	SNR Depois (dB)
Block Thresholding	17,44	27,39
Spectral Subtraction	17,44	9,47
Wavelet Thresholding	17,44	27,61

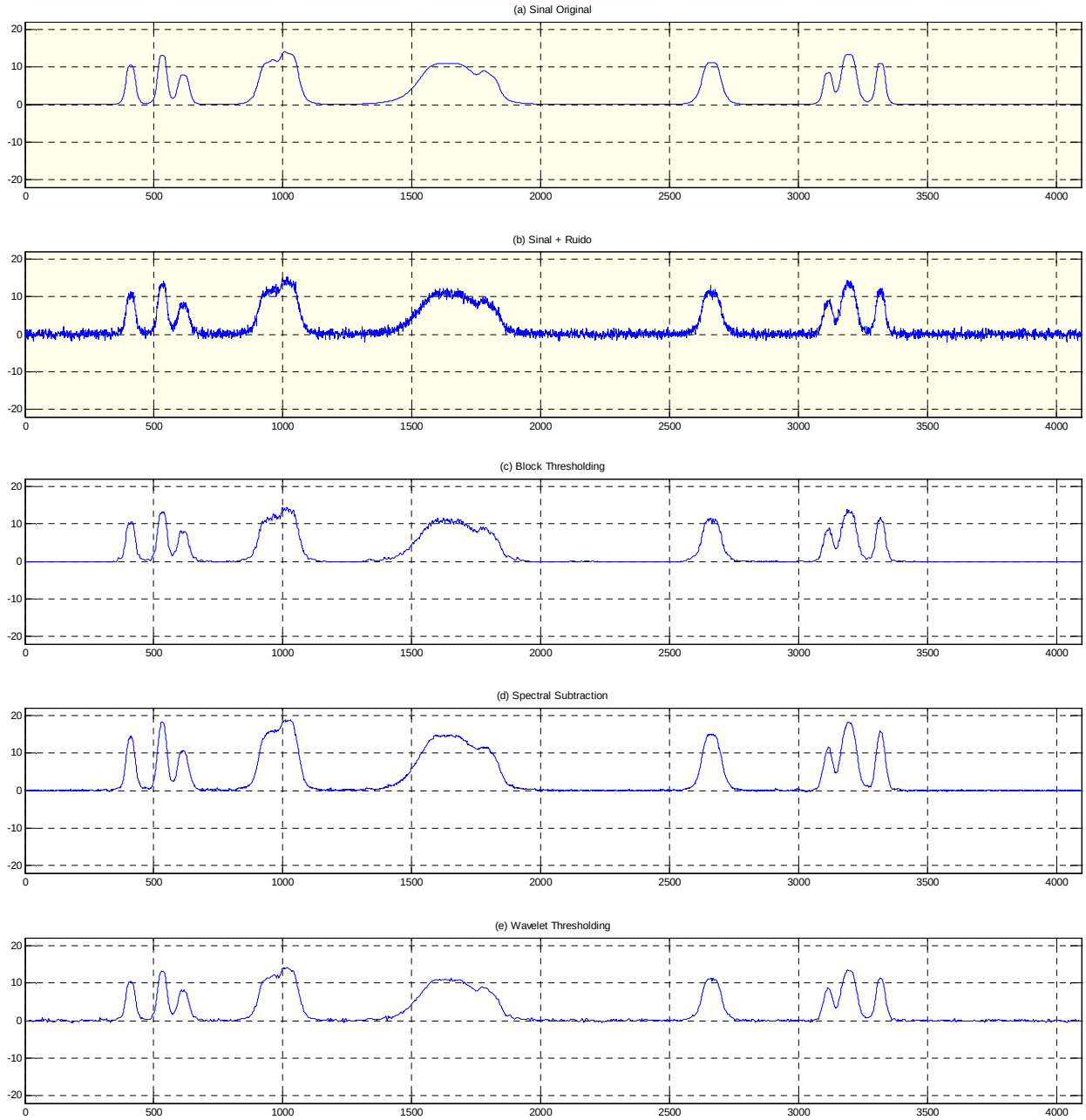

Figura 4.13 (a), (b): *Bumps* original e ruidosa. (c), (d), (e): *Bumps* tratada com cada uma das técnicas indicadas. O eixo vertical representa a magnitude e o horizontal o número de amostras pontuais.

4.1.1.3 *Heavy Sine*

Heavy Sine é uma função que mescla um seno bastante suave, com transições abruptas sutis. Essas características põem à prova as técnicas. *Wavelet Thresholding* possui o melhor desempenho, atestado pelo SNR (Tabela 4.4), inclusive. *Block Thresholding* também executa um bom trabalho. *Spectral Subtraction* não vai bem e o motivo é simples: não há amostras pontuais contendo apenas ruído no trecho inicial, e isso é um pré-requisito para o uso da técnica. Esse fato foi propositalmente deixado neste teste para exibir isso. Além de praticamente não remover o ruído, pode-se observar que ela distorce o sinal nas regiões com as transições abruptas (Figura 4.14(d)).

Tabela 4.4 SNR antes/depois da filtragem da função *heavy sine*

Técnica	SNR Antes (dB)	SNR Depois (dB)
Block Thresholding	16,60	19,13
Spectral Subtraction	16,60	9,38
Wavelet Thresholding	16,60	27,38

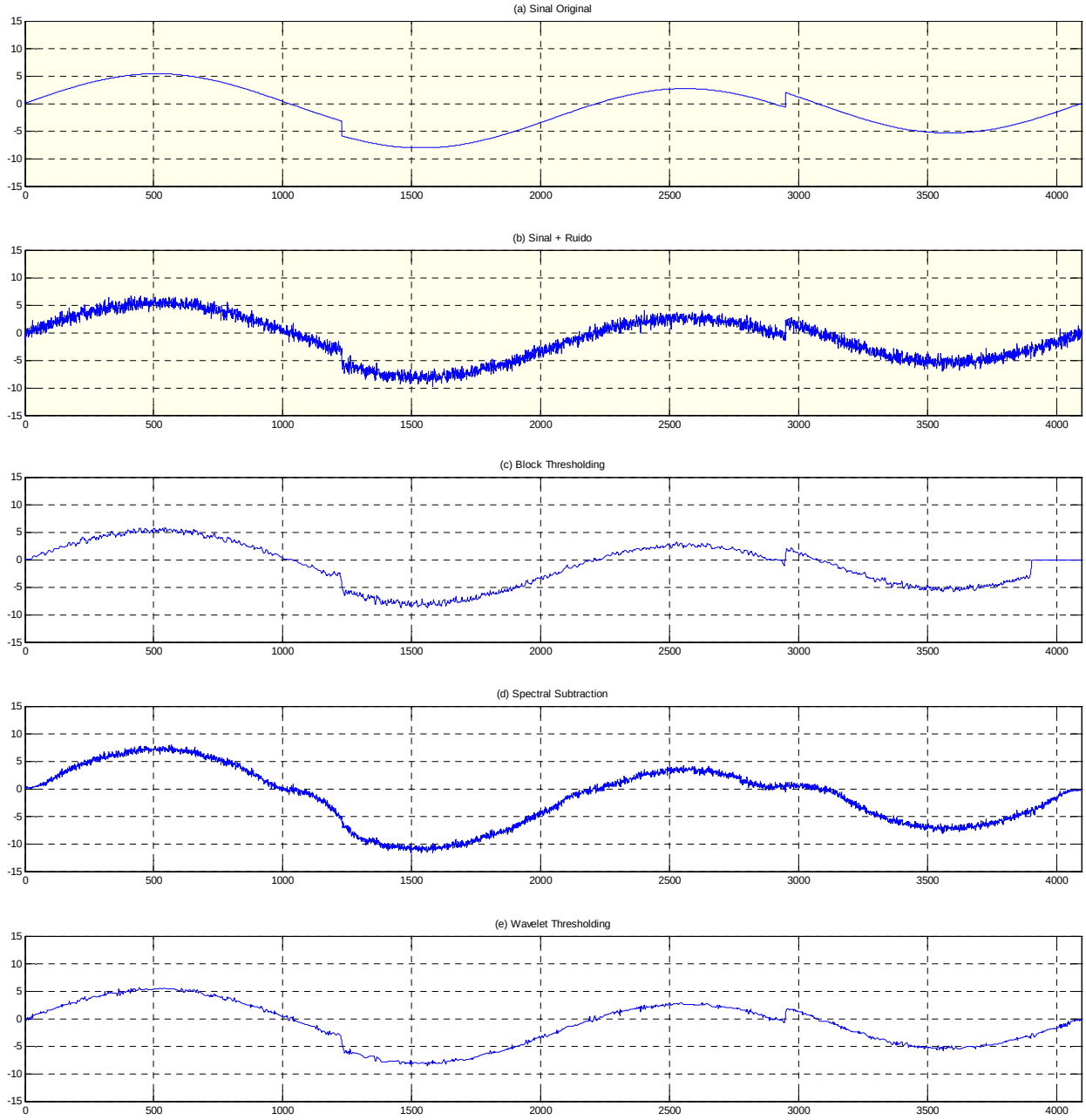

Figura 4.14 (a), (b): *Heavy Sine* original e ruidosa. (c), (d), (e): *Heavy Sine* tratada com cada uma das técnicas indicadas. O eixo vertical representa a magnitude e o horizontal o número de amostras pontuais.

4.1.1.4 *Doppler*

Doppler é uma função que gera um sinal que exibe um comportamento contínuo, mas com transições mais rápidas sendo progressivamente suavizadas. *Wavelet Thresholding*, *Block Thresholding* e *Spectral Subtraction*, podem ser ordenadas em ordem decrescente de eficácia neste exemplo (Tabela 4.5). *Wavelet Thresholding* limpou bastante o sinal, enquanto que a *Spectral Subtraction* deixou mais ruído. Motivo? Também não há amostras pontuais exclusivamente ruidosas suficientes no começo do sinal para a aplicação da técnica. Ainda

assim, surpreendemente, o sinal filtrado é, visualmente, melhor que o esperado (Figura 4.15(d)).

Tabela 4.5 SNR antes/depois da filtragem da função *doppler*

Técnica	SNR Antes (dB)	SNR Depois (dB)
Block Thresholding	16,19	23,29
Spectral Subtraction	16,19	9,26
Wavelet Thresholding	16,19	26,87

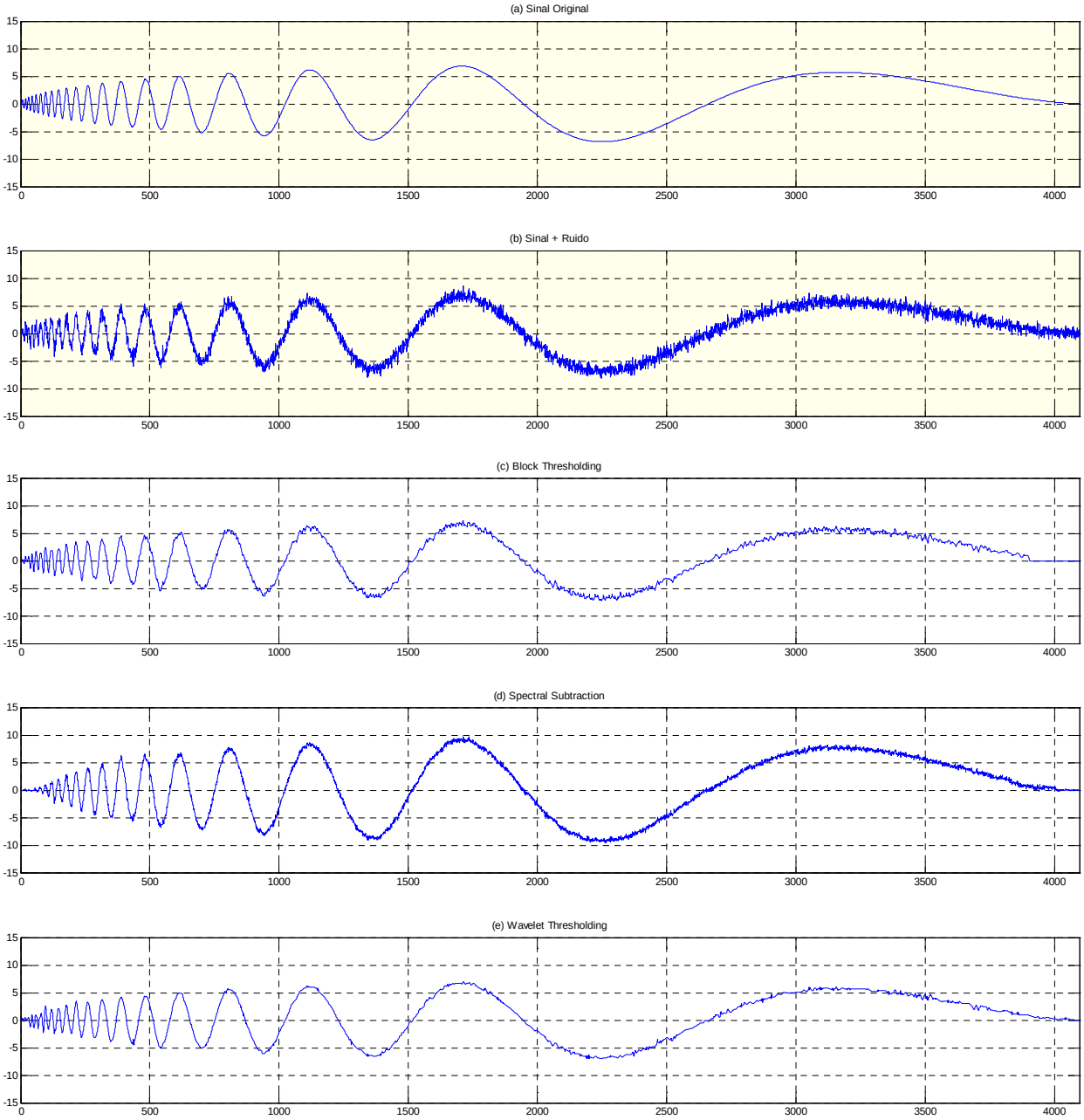

Figura 4.15 (a), (b): *Doppler* original e ruidosa. (c), (d), (e): *Doppler* tratada com cada uma das técnicas indicadas. O eixo vertical representa a magnitude e o horizontal o número de amostras pontuais.

4.1.1.5 *Quadchirp*

Esta é uma função que gera um sinal sintético complicado. Ao contrário da anterior, inicia suas oscilações com baixa frequência, mas rapidamente atinge frequências muito elevadas. A avaliação visual (Figura 4.16) é muito difícil neste caso, mas a julgar pelos resultados do SNR (Tabela 4.6), nenhuma técnica teve desempenho satisfatório.

Tabela 4.6 SNR antes/depois da filtragem da função *quadchirp*

Técnica	SNR Antes (dB)	SNR Depois (dB)
Block Thresholding	15,99	12,10
Spectral Subtraction	15,99	9,76
Wavelet Thresholding	15,99	14,48

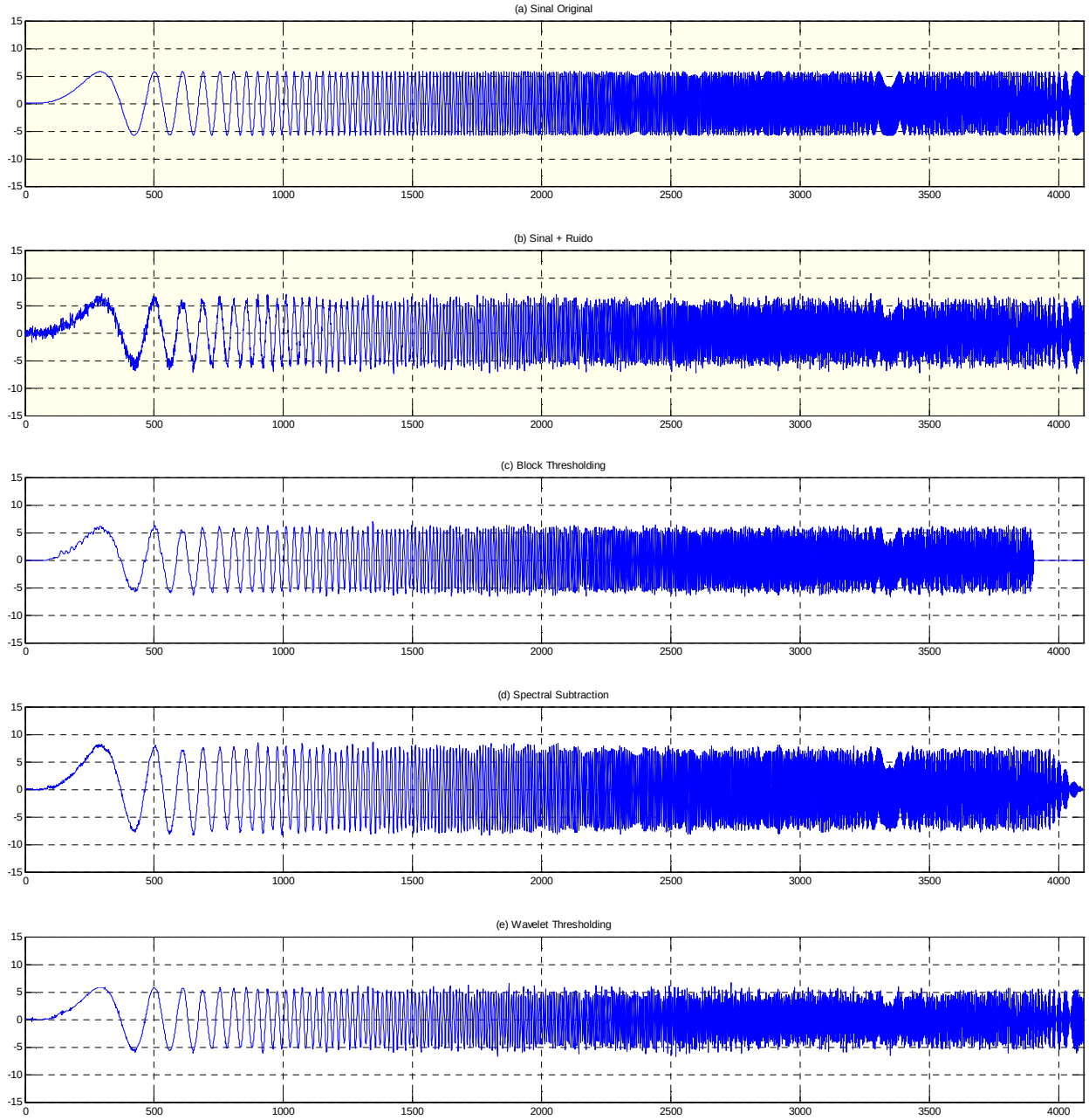

Figura 4.16 (a), (b): *Quadchirp* original e ruidosa. (c), (d), (e): *Quadchirp* tratada com cada uma das técnicas indicadas. O eixo vertical representa a magnitude e o horizontal o número de amostras pontuais.

4.1.1.6 *Mishmash*

Mishmash é, como o próprio nome sugere, uma completa confusão/desordem. Todas as técnicas tiveram péssimo desempenho (Tabela 4.7), em especial a *Wavelet Thresholding* que, em virtude de usar os coeficientes de detalhe do primeiro nível de decomposição, certamente selecionou um *threshold* bem elevado (em virtude da componente aleatória do sinal), o que justifica o sinal ter sido completamente danificado no processo de filtragem (Figura 4.17).

Tabela 4.7 SNR antes/depois da filtragem da função *mishmash*

Técnica	SNR Antes (dB)	SNR Depois (dB)
Block Thresholding	15,99	11,90
Spectral Subtraction	15,99	6,24
Wavelet Thresholding	15,99	1,25

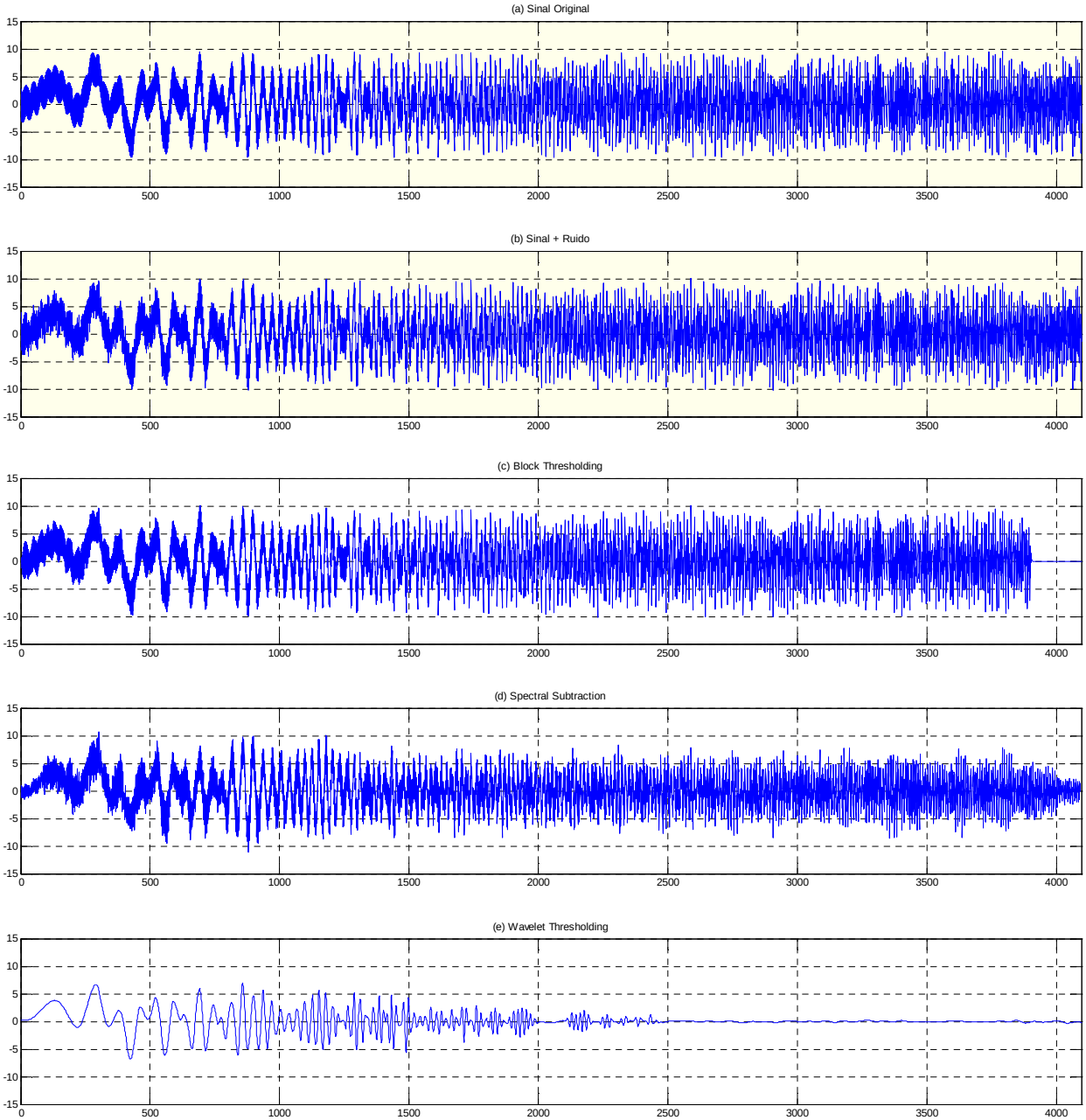

Figura 4.17 (a), (b): *Mishmash* original e ruidosa. (c), (d), (e): *Mishmash* tratada com cada uma das técnicas indicadas. O eixo vertical representa a magnitude e o horizontal o número de amostras pontuais.

4.1.2 Comentários Gerais dos Testes com Sinais Sintéticos

Ainda são um tanto escassos os materiais específicos sobre remoção de ruído em sinais de áudio musical. Há muitas pesquisas envolvendo sinais de voz e sinais unidimensionais "não categorizados" (ou seja, não específicos). No inicio dessa dissertação, deparou-se com a seguinte indagação: sinais de áudio são sinais 1D discretos, então qual seria seu comportamento frente às técnicas de remoção de ruído descritas na literatura como muito boas, exibindo excelentes resultados em sinais sintéticos (também unidimensionais)?

Conforme demonstrado nos testes dessa sessão, é indiscutível o ótimo desempenho de técnicas modernas, como o *Wavelet Thresholding*, em sinais unidimensionais. Mas conforme será mostrado em sessões posteriores, com sinais reais (músicas), é falsa a ideia de que uma técnica com bom desempenho com sinais sintéticos obrigatoriamente manterá esta qualidade com outros tipos de sinais, mesmo que também unidimensionais (como é o caso).

4.2 *Denoising* de Sinais Reais (Músicas)

Após os testes sintéticos, procederam-se os testes com amostras reais, no caso, trechos de músicas contendo apenas áudio instrumental (sem voz). Inicialmente, serão melhor caracterizados alguns termos metafóricos usados em análises subjetivas de áudio. A seguir, cada uma das amostras será detalhadamente comentada, o que servirá de base de comparação para os comentários pós-filtragem. Uma sessão inteira avaliará o desempenho das técnicas testadas de diferentes formas: objetiva, subjetiva e computacionalmente.

4.2.1 Algumas Definições De Termos Metafóricos Usados Em Análises Subjetivas De Áudio

Um efeito colateral comum encontrado nos sinais de áudio pós-filtragem (para remoção de ruído) é como muitos autores descrevem como uma aparência "lavada" (*washed*), ou que o som perde (total ou parcialmente) o seu "brilho". Isto se dá porque esse "brilho" é dado por estruturas tempo-frequência muito sutis, de magnitude próxima ou inferior à do ruído, concentradas especialmente em frequências mais altas. Do ponto de vista físico, muitas delas são de fato os harmônicos das notas dos instrumentos. Por exemplo, uma nota Lá padrão, num instrumento de cordas como um violão, tem sua frequência fundamental (consequentemente a de maior magnitude) centrada em 440 Hz. Contudo, o som emitido pelo instrumento não está limitado apenas a essa frequência. Simultaneamente, a corda emite diversos outros harmônicos representados por frequências múltiplas da fundamental, mas com

intensidade significativamente menor. E sim, nós somos capazes de perceber isso, especialmente a ausência disso. A nota fundamental, por ter maior magnitude, é mantida, enquanto a maior parte dos harmônicos é removida.

O fato comentado acima pode ser visualizado nos espectrogramas exibidos na Figura 4.18. Neste trecho, há um violão que é o responsável por esses ataques exibidos no espectrograma superior. A nota fundamental de cada acorde tocado é de frequência mais baixa e magnitude maior (em tons laranja/vermelho). Mas o som do acorde como um todo é dado por diversas outras frequências (harmônicos) que são emitidos simultaneamente, com frequência mais alta e de magnitude menor (em tons verde/amarelado). O ponto em questão é que essas estruturas que dão o "brilho" ao som do violão praticamente desaparecem, como pode ser observado no espectrograma inferior da Figura 4.18, que corresponde ao sinal processado com a técnica *Block Thresholding*. Permanecem, no entanto, as notas fundamentais, de maior magnitude.

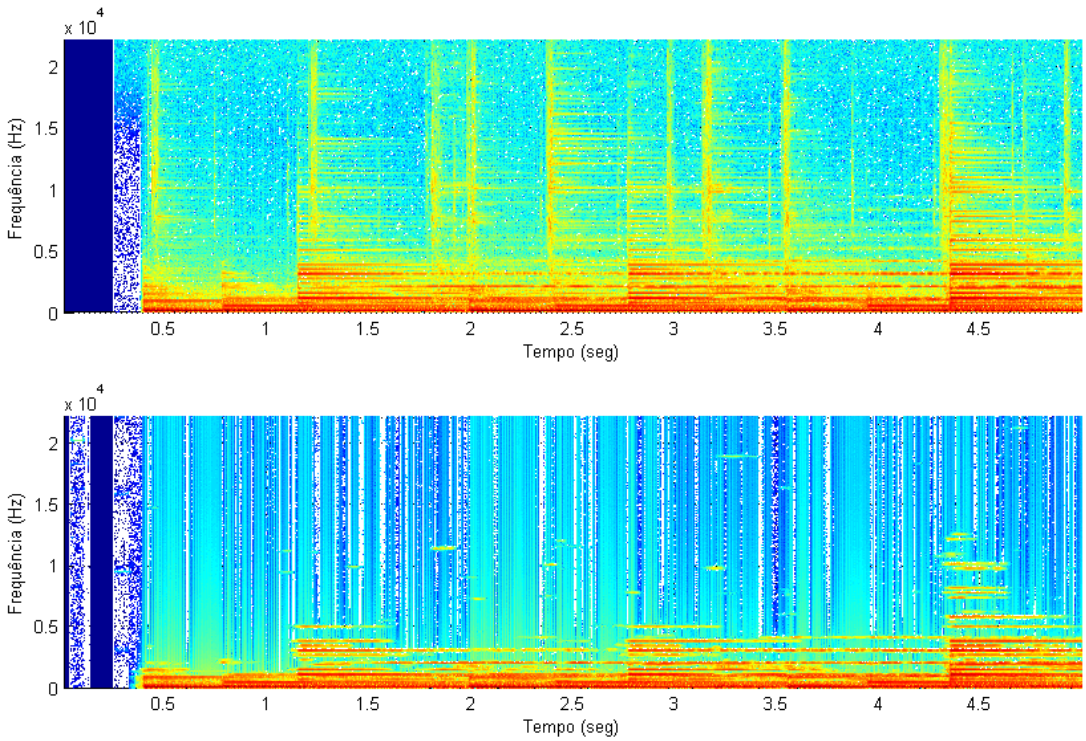

Figura 4.18 Amostra: *The Verve - The Drugs Don't Work*. Espectrogramas: original (acima) e pós remoção do ruído (abaixo)

Descrever textualmente aspectos relacionados a sinais de áudio é uma tarefa um tanto complicada. Por isso, outra forma de entender o fenômeno descrito acima é fazendo uma analogia com o processamento de imagens. Tome, por exemplo, a tradicional imagem "baboon", exibida na Figura 4.19. No topo tem-se a imagem inalterada. Ao meio, a mesma

contaminada com um ruído gaussiano monocromático. E, abaixo, a simulação de um processo de filtragem (para a remoção do ruído). É possível observar nitidamente que a imagem filtrada possui pouquíssimo ruído. Porém, a maioria dos detalhes da imagem original também se foi. A imagem “base” continua reconhecível, mas possui uma aparência “lavada”, em relação à imagem original.

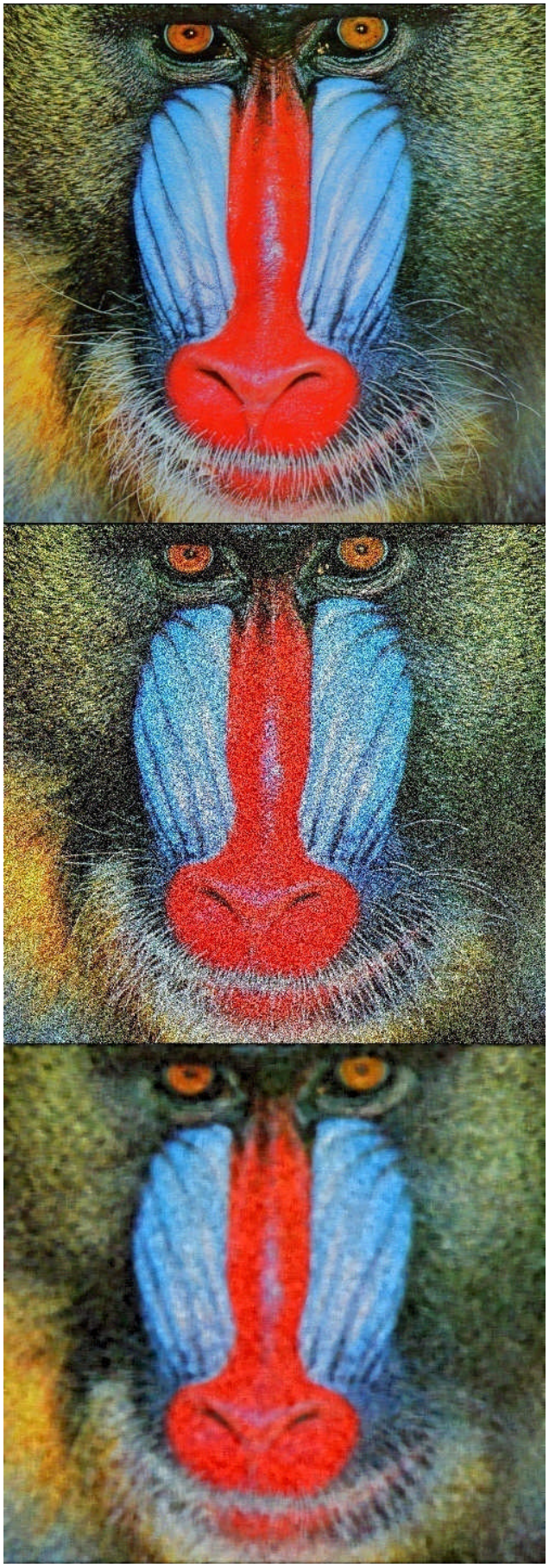

Figura 4.19 Imagem Original (Topo), Ruidosa (Meio) e Filtrada (Inferior)

4.2.2 Caracterizando as Amostras Originais

Do ponto de vista técnico, todas as amostras foram extraídas direto do áudio original do CD: 44100 Hz/16 bits/*Stereo*. As únicas alterações foram os cortes (a fim de "enquadrar" a região de interesse) e a escolha de apenas um canal, tornando a amostra *mono* (existência de apenas um canal de áudio). É importante frisar que nenhuma das amostras originou-se de arquivos previamente comprimidos (com perdas), como o MP3. Isso poderia prejudicar as análises em virtude do acúmulo de distorções.

A Tabela 4.8 sumariza algumas informações sobre as amostras.

Tabela 4.8 Amostras usadas nos testes

Título	**Artista**	**Duração (em min.)**	***Samples***
Life In Technicolor II	*Coldplay*	1:12.148	3.181.746
Square One	*Coldplay*	0:16.230	715.740
Somewhere Only We Know	*Keane*	0:23.021	1.015.234
Untitled 1	*Keane*	0:45.699	2.015.333
What I've Done	*Linkin Park*	0:40.310	1.777.677
Apologize	*OneRepublic*	0:32.417	1.429.582
Bitter Sweet Symphony	*The Verve*	1:06.981	2.953.848
Sonnet	*The Verve*	0:21.887	965.225
The Drugs Don't Work	*The Verve*	0:24.079	1.061.867
Flute Concerto	*Vivaldi*	0:23.789	1.049.091

Cada uma das amostras foi cuidadosamente escolhida a fim de possuir alguns elementos importantes à análise da eficácia de cada técnica.

A seguir, cada amostra é comentada. Aspectos musicais e técnicos são abordados, os quais espera-se que serão úteis na descrição da forma como cada técnica de remoção de ruídos afeta o sinal original. Os subtrechos, quando notados, estão cotados em segundos.

Coldplay – Life In Technicolor II

Trecho da primeira faixa do EP *Prospekt's March* da banda inglesa *Coldplay*, essa amostra progressivamente adiciona instrumentos de modo que a complexidade do áudio vai aumentando. Abaixo, cada subtrecho é comentado:

- [0 – 16] Inicia com um som sintetizado que lembra o de um berimbau. O ritmo é razoavelmente marcado[12].
- [16 – 40] É adicionado mais um instrumento: violão com cordas de aço (som mais metalizado que das cordas de *nylon* comuns). Além disso, é possível observar outro som sintetizado predominantemente de baixa frequência, com timbre de instrumento de sopro.
- [40 – final] A bateria entra progressivamente e reforça consideravelmente a marcação do ritmo e a ocupação do espectro de frequências.

Por que esta amostra foi escolhida? Um dos motivos é pelo fato da complexidade aumentar progressivamente. Como a marcação do ritmo varia, é possível observar até que ponto a filtragem prejudicará os elementos transientes do sinal. E se, no último subtrecho, com o espectro mais "preenchido" seria possível continuar observando mais ou menos distorções.

Coldplay – Square One

Mais uma amostra extraída de uma música da banda *Coldplay*. Do terceiro álbum em estúdio, esta faixa tem características completamente diferentes da citada anteriormente. Pode ser avaliada como tendo apenas um trecho, contínuo, abaixo comentado:

- [trecho único] Os sons de maior intensidade são predominantemente de baixa frequência, provavelmente provenientes de um teclado, mas cujas notas variam de maneira muito suave. Não há qualquer elemento transiente nesta amostra. Em outras palavras, o ritmo não é marcado. Contudo, há um aspecto muitíssimo interessante e útil: há harmônicos de baixa intensidade ao longo de quase todo o espectro. Além de serem visíveis no espectrograma, são também bastante audíveis e responsáveis pelo "brilho" da música.

Por que esta amostra foi escolhida? O motivo, já pontuado acima, é a ausência de marcação de ritmo, predomínio de baixa frequência, mas com harmônicos de baixa intensidade ao longo de quase todo o espectro. Esses harmônicos se mostrarão importantes elementos de análise nas sessões seguintes.

[12] Tecnicamente falando, o ritmo é tanto mais "marcado" quanto mais evidentes forem os picos observados no espectrograma do sinal. Geralmente o instrumento que mais "marca" o ritmo são os percursivos, como a bateria. Do ponto de vista de sinais, tais trechos são chamados de **transientes**, e se caracterizam por ocupar uma faixa de frequência muito ampla (geralmente todo o espectro do sinal) durante períodos curtos, quase impulsivos. Devido à extensa faixa espectral ocupada, tais sons são importantes na caracterização do desempenho das técnicas, já que estes costumam evidenciar mais claramente as distorções provocadas pela filtragem.

Keane – Somewhere Only We Know

Trecho inicial da primeira faixa, do primeiro álbum da banda inglesa *Keane*, esta faixa é abaixo analisada:

- [trecho único] Dentre todas as amostras, esta talvez seja a que possua o espectro mais "cheio". O ritmo é fortemente marcado pela bateria, que acompanha um piano.

Por que esta amostra foi escolhida? O objetivo é investigar como percebemos as distorções inseridas pela filtragem do ruído em situações onde o espectro é muito amplo e intensamente preenchido (potência do áudio original é alta em relação ao ruído).

Keane – Untitled 1

Amostra advinda da décima faixa, do mesmo álbum do *Keane*, comentado anteriormente, esta é uma música com ritmo fortemente marcado, conforme comentado abaixo:

- [0 – 9] Prevalece um ritmo percussivo ditado por uma bateria eletrônica. Pelo espectrograma é possível observar transientes uns mais curtos que outros, porém marcadas de maneira impecavelmente regular. Não há predomínio de baixa ou alta frequência. Os pulsos se distribuem de maneira razoavelmente uniforme ao longo do espectro.
- [9 – 28] O ritmo percussivo sem mantém, mas com a presença de um teclado inserindo elementos harmônicos (notas com transição suave) ao fundo.
- [28 – final] Mais um conjunto de notas de baixa frequência completa a harmonia da música.

Por que esta amostra foi escolhida? Devido ao seu ritmo percussivo muito bem marcado (intensidade e regularidade) e a presença simultânea de harmônicos de baixa frequência.

Linkin Park – What I've Done

Amostra extraída do trecho inicial da faixa "*What I've Done*" do álbum "*Minutes To Midnight*" da banda de *punk-rock* americana *Linkin Park*, esta é uma amostra com características interessantes, comentadas abaixo:

- [0 – 3] Começa com um som sintetizado de baixa/média frequência que muito se assemelha a um ruído.
- [3 – 7] Este subtrecho contém exclusivamente algumas notas tocadas por um teclado.

- [7 – 16] O teclado permanece, porém uma bateria eletrônica agora o acompanha. O ritmo é bem marcado e muito regular, porém não se estendendo ao longo de todo o espectro de frequência.
- [16 – 22] A percussão eletrônica ganha mais força. Já se observam distorções (na forma de ruído) propositalmente adicionadas pela banda.
- [22 – final] Aqui encontra-se o motivo principal que motivou a escolha dessa amostra. O ritmo percussivo se mantém em plano de fundo, uma vez que em primeiro plano é possível ouvir uma guitarra bastante distorcida que, juntamente com os pratos da bateria (som metálico), preenche praticamente todo o espectro audível.

Por que esta amostra foi escolhida? Devido ao seu último subtrecho, comentado acima. O alto grau de distorções que se assemelham muito a um ruído que naturalmente faz parte da música motiva as perguntas: Conseguiríamos perceber de maneira muito evidente ruídos adicionais inseridos nesse trecho? Ou ainda, como a técnica iria se comportar num trecho onde ruído e música se misturam tanto?

OneRepublic – Apologize

Trecho extraído da faixa *Apologize*, do álbum "*Dreaming Out Load*" da banda de *pop-rock* americana *OneRepublic*, abaixo comentada:

- [0 – 8] Inicialmente há uma harmonia com uma espécie de violão cello/baixo sintetizado. O ritmo é muito levemente marcado, com picos não muito intensos no espectrograma.
- [8 – 16] É adicionada a bateria que marca pontualmente, gerando transientes intensos seguidos de sons que imitam castanholas, instrumento percussivo que gera perturbações muito nítidas e rápidas no sinal.
- [16 – final] A harmonia do primeiro trecho é agora mais evidente, ocupando uma porção maior do espectro e, junto às castanholas do trecho anterior, contribui para criar uma espacialidade (como se desse para sentir o ambiente em que a música está sendo produzida, através das reverberações) característica dessa música.

Por que esta amostra foi escolhida? Os transientes fortes e muito rápidos criados pelas castanholas e a espacialidade que o som logo em seguida carrega, certamente estará "pondo à prova" qualquer que seja a técnica de remoção de ruídos aplicada.

The Verve - Bitter Sweet Symphony

Conhecidíssima música da banda inglesa *The Verve*, esta será uma amostra muito importante aos testes em virtude de alguns elementos, comentados abaixo:

- [0 – 12] A faixa começa com uma harmonia dada por um instrumento de cordas onde prevalecem notas mais baixas, ocupando apenas a parte mais baixa do espectrograma.
- [12 – 46] Progressivamente um violino vai produzindo uma sequência de notas, porém com ritmo bem marcado, produzindo estruturas em quase todas as frequências do espectro (harmônicos). A espacialidade observada pela reverberação das notas do violino é notável.
- [46 – final] A bateria chega marcando muito bem o ritmo, gerando transientes que se misturam a todos os harmônicos dos outros instrumentos.

Por que esta amostra foi escolhida? A beleza dessa música está na simplicidade artística em meio a certa complexidade do ponto de vista de processamento de sinais. A presença de longos harmônicos, junto com transientes curtos e de baixa magnitude impõem um grau de dificuldade enorme ao bom desempenho de qualquer técnica.

The Verve - Sonnet

Trecho inicial da faixa *Sonnet*, do álbum "*Urban Hymns*" de 1997, esta é mais uma amostra com características úteis. Seus subtrechos são comentados abaixo:

- [0 – 10] A princípio são tocados alguns acordes com um violão com cordas de aço. As notas baixas são nítidas tanto no espectrograma quanto pelo som ouvido. O timbre metálico (em virtude das cordas serem de metal) das cordas do violão ocupa todo o espectro de frequência. O ritmo é bem marcado, assim, esse trecho é quase que inteiramente composto por transientes curtos e algumas vezes bem sutis (baixa intensidade).
- [10 – final] O violão permanece, porém o ritmo agora é mais fortemente marcado pela bateria, reforçando os transientes. Uma guitarra emite algumas notas curtas no ritmo ditado pela bateria. Há também algumas notas harmônicas ao fundo, ditadas por um baixo elétrico.

Por que esta amostra foi escolhida? O som naturalmente metálico das cordas de aço do violão que inicia essa música tem um "brilho" notável. Tecnicamente falando, esse brilho é representado por componentes frequenciais que ocupam praticamente todo o espectro do

sinal. E aí é que está o ponto – se essas componentes de mais alta frequência do sinal forem suprimidas na filtragem, esse brilho irá desaparecer, parcial ou totalmente.

The Verve - The Drugs Don't Work

Mais um trecho de uma faixa do mesmo álbum citado na amostra anterior, é abaixo comentado:

- [trecho único] Há três instrumentos que dão o tom dessa faixa: dois violões e um teclado (fazendo um som de timbre semelhante ao de um violino). Dos dois violões, um é dedilhado e o outro, com cordas de metal, produz acordes que formam a base da música. O teclado emite harmônicos longos.

Por que esta amostra foi escolhida? O violão que faz os acordes da base da música tem um brilho semelhante ao já comentado na outra faixa da mesma banda, porém nesta amostra ele está muito mais sutil. Ele acaba por criar várias estruturas tempo-frequência de baixa magnitude, mas muito importantes à apreciação da canção. São muitos transientes, porém de baixa intensidade. Provavelmente será muito difícil remover ruído em meio a esses elementos citados.

Vivaldi - Flute Concerto

Única amostra proveniente de uma música clássica, um concerto para flauta do italiano *Vivaldi*, é abaixo comentada:

- [trecho único] O instrumento principal é uma flauta que soa muito aguda. Suas notas ocupam trechos muito bem definidos do espectro, além de emitirem harmônicos muito evidentes. Se fosse só por ela, o espectro seria bem heterogêneo, mas não é pois há também um cravo fazendo a base da música. Esse último emite um som metalizado (típico das cordas de metal desse instrumento) que preenche boa parte do espectro e dá brilho à musica.

Por que esta amostra foi escolhida? A regularidade das notas da flauta é interessante do ponto de vista da filtragem. É de se esperar que essas estruturas sejam mantidas, dado que são bem marcantes. Já não se pode esperar o mesmo do brilho que vem do som metálico do cravo.

4.2.3 Adicionando o Ruído

A fim de simular a presença de ruído nas amostras, adicionou-se um ruído branco gaussiano aditivo de média zero e variância 0,0001. Tal variância corresponde a 0,01% da máxima amplitude possível do sinal de áudio das amostras, que está na escala [-1;+1] em

valores de ponto flutuante. Estes valores, obtidos experimentalmente, foram adotados por gerarem um ruído considerável e bastante perceptível. Esse valor de ruído é suficiente para sobrepor algumas partes das amostras. Em outras palavras, o ruído é mais intenso que o sinal original em alguns trechos, o que levará a algumas conclusões, abordadas adiante.

4.2.3.1 A Audição Humana em Meio ao Ruído

A natureza nos dotou de uma capacidade impressionante de isolar estruturas/padrões mesmo em meios ruidosos. Isso é notável não apenas no que diz respeito à visão, mas também quando se tratam de sons. Numa orquestra com dezenas de instrumentos, um ouvinte treinado consegue "separar" mentalmente cada um deles. Mesmo ouvintes não treinados usam dessa mesma capacidade quando conseguem estabelecer diálogo em meio a uma multidão, com trânsito de carros, obras, e demais ruídos.

Ao ouvir atentamente as amostras contaminadas por ruído, apesar de se poder perceber a presença marcante do ruído, ainda assim foi possível distinguir entre elementos originais da música e o que era ruído. O fato é que muitos desses elementos desaparecem após a filtragem, especialmente no caso dos mais sutis. De fato, em parte é possível acreditar na hipótese do cérebro reconhecer tais estruturas devido ao conhecimento prévio do áudio original. Portanto, são apenas conjecturas que precisariam de uma abordagem cientifica a fim de serem provadas.

4.2.4 Removendo o Ruído

Nesta sessão, inicialmente serão explicitados os principais parâmetros de configuração de cada técnica usada nos testes. Em seguida, feitos comentários subjetivos de cada amostra tratada, antes de serem efetuadas as avaliações formais.

4.2.4.2 Configuração dos Parâmetros de Cada Técnica

Cada técnica testada neste trabalho carrega suas próprias particularidades, conforme toda a abordagem teórica do Capítulo 3. Os parâmetros usados nos testes de cada uma delas são especificados abaixo.

Time-Frequency Block Thresholding

- Tamanho da janela: 50 ms
- Desvio padrão do ruído: 0,01
- Taxa de amostragem: 44.100 Hz

Spectral Subtraction

- Silêncio inicial: 0,25 s

- Taxa de amostragem: 44.100 Hz
- α: 10
- β: 0,01

Wavelet Thresholding

- Número de níveis: 10
- Regra de seleção do threshold: *SURE*
- Método de *thresholding*: *soft thresholding*
- Reescalamento dos coeficientes: SNL (estimação única baseada no primeiro nível)
- Família de wavelet: *bior6.8*

4.2.4.3 Amostras Tratadas: Comentários Subjetivos

O objetivo dessa sessão é, antes de proceder as demais análises, comentar de maneira mais subjetiva e informal as experiências e resultados obtidos com cada técnica sobre cada amostra, individualmente. Isto porque, dadas as particularidades de cada amostra (conforme comentado detalhadamente na sessão anterior), era de se esperar que os resultados diferissem. Neste capítulo serão comentadas amostra por amostra, e ficará a cargo do capítulo seguinte uma análise mais conclusiva sobre o desempenho geral de cada técnica. Mantendo a coerência com a forma como cada amostra original foi comentada, as amostras pós-filtragem também serão comentadas por subtrechos, denotados em segundos.

A fim de simplificar a escrita, as técnicas serão referenciadas por siglas, a saber:

- **TFBT**: *Time-Frequency Block Thresholding*
- **SS**: *Spectral Subtraction*
- **WT**: *Wavelet Thresholding*

Coldplay – Life In Technicolor II

- [0 – 16] Parte da "espacialidade" presente no primeiro trecho dessa amostra deve-se às diversas componentes frequenciais mais altas, mas que, por possuírem uma intensidade menor, foram em sua grande parte limadas pelas técnicas TFBT e SS. Além disso, em virtude dos elementos transientes serem curtos e variarem bastante em pouco tempo, ambas as técnicas tenderam a "sujar" o áudio de maneira perceptível, prejudicando a "clareza" das notas. É digno de nota, porém, que apesar de ser possível observar em ambas problemas parecidos, a técnica TFBT foi muito mais eficiente na remoção do ruído, sem falar na inexistência do ruído musical, que, apesar de reduzido, ainda é possível observar na técnica SS. Porém, à medida que a técnica TFBT remove

mais ruído, deixa o som artificialmente mais "metalizado". A técnica WT não teve problemas em manter as componentes mais altas, porém o desempenho geral foi muito ruim, deixando muito ruído e ainda inserindo mais alguns artefatos inexistentes na amostra original.

- [16 – 40] Com o ritmo mais fortemente marcado, a intensidade das componentes mais altas também aumentou, permitindo uma distinção mais clara entre ruído e sinal original, favorecendo o desempenho das técnicas TFBT e SS. TFBT ainda dá resultados muito melhores nas transientes, sujando (misturando) muito menos os sons de ataque. Esse melhor desempenho pode ser atribuído à capacidade adaptativa da técnica de encontrar o tamanho do bloco mais apropriado. Ainda que haja algum ruído musical na SS, a percepção auditiva dele é muito mais baixa, uma vez que o som é mais complexo que no primeiro trecho. Porém, é possível observá-lo graficamente no espectrograma. Quanto aos harmônicos de baixa frequência, ambas as técnicas preservaram. WT manteve seu desempenho ruim do primeiro trecho, removendo muito pouco do ruído.
- [40 – final] Neste último subtrecho o áudio é bem mais preenchido, tornando menos perceptível o ruído, mesmo no sinal ruidoso, sem tratamento. Entre TFBT e SS a diferença diminui, porém TFBT remove bem mais o ruído. WT continua com desempenho ruim.

Coldplay – Square One

- [trecho único] Esta é uma amostra bastante simples, conforme já comentado. Contendo apenas harmônicos contínuos e longos, e a ausência de transientes, era de se esperar que esta fosse uma amostra fácil de remover o ruído. Contudo, apesar de predominarem as baixas frequências, há notáveis frequências médias e altas que dão o brilho que a música possui neste trecho. E, como imaginado, todas as técnicas falham em mantê-lo. Tanto TFBT quando SS deixam o som "lavado" (sem brilho). Era de se esperar que isso acontecesse, uma vez que as estruturas que dão esse brilho ao som tem intensidade próxima ou menor que o ruído, ficando assim sobrepostas. Com a SS é ainda possível notar uma clara presença de ruído musical ao longo de toda a amostra. Isto pode ser observado no espectrograma inferior da Figura 4.20. Uma comparação com o espectrograma superior dessa mesma figura (da técnica TFBT) revela que a técnica TFBT consegue de fato evitar o ruído musical presente na SS (visualmente representado pelos tons amarelos sobre o fundo verde). É possível observar ainda a

regularização tempo-frequência, denotada pelo espectro mais "regular" nas frequências de baixa magnitude (tons azul/verde). WT teve um desempenho razoável, removendo praticamente todo o ruído nas frequências médias e altas. Mas ainda restou muito ruído na faixa de frequência que concentrava a maior parte da informação (baixa frequência). Sem falar que o áudio soa como se estivesse sendo emitido por um antigo rádio AM (devido, principalmente, a baixa extensão espectral).

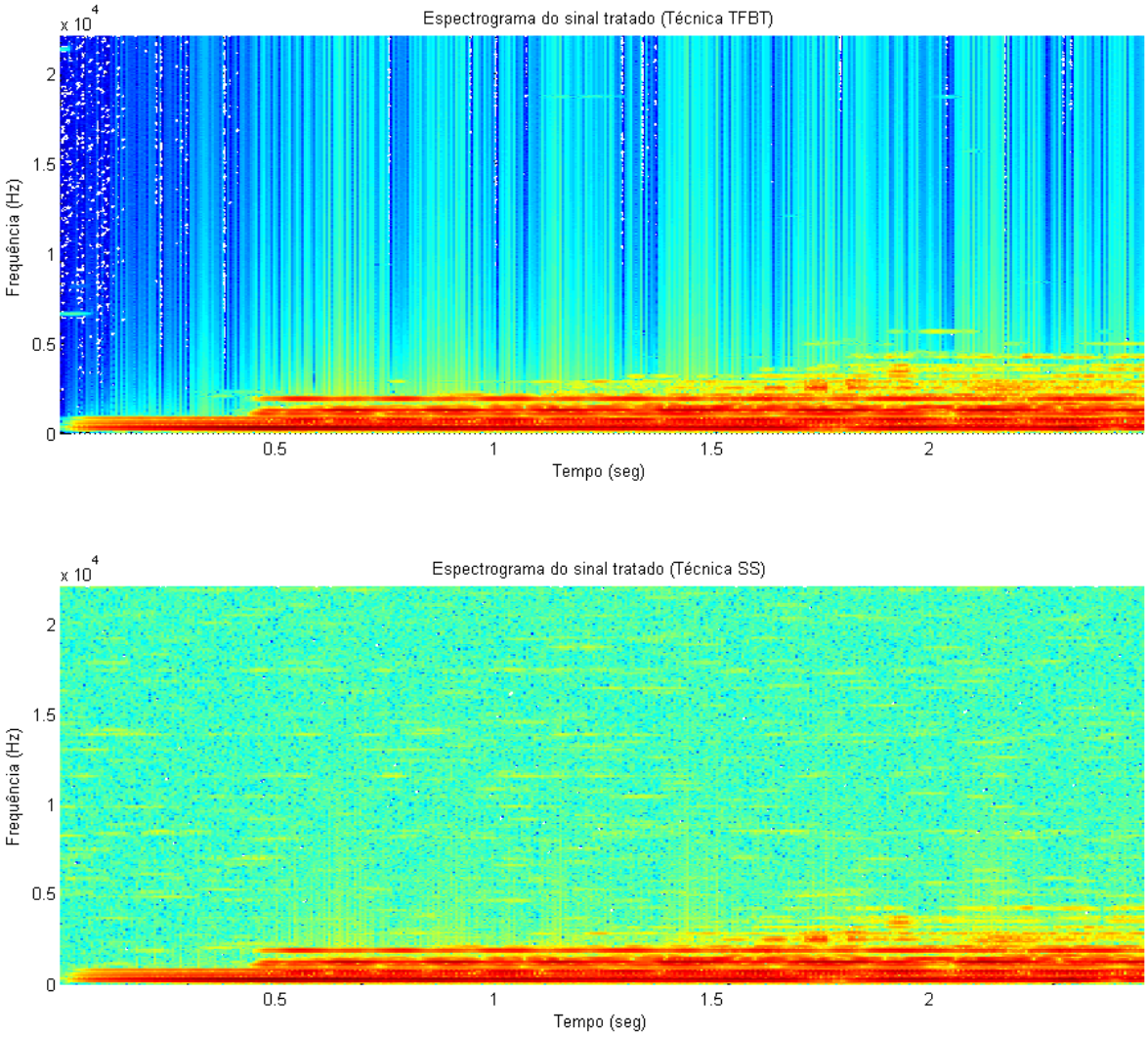

Figura 4.20 Espectrogramas da amostra "*Coldplay - Square One*" tratados com TFBT (superior) e SS (inferior). É nítida a presença de ruído musical (denotado em tons amarelos sobre fundo verde) na imagem inferior, em contraste com a aparencia regular da imagem superior.

Keane – Somewhere Only We Know

- [trecho único] O espectro "cheio" e com quase todas as frequências com magnitude média a alta, torna mais difícil a percepção do ruído residual ou das distorções inseridas pelas técnicas de processamento. Apesar de algumas diferenças visíveis nos espectrogramas, o sinal tratado com TFBT e SS soam de maneira bem parecida. A

menos dos artefatos já observados em outras amostras, WT é suficientemente bom nesta amostra.

Keane – Untitled 1

- [0 – 9] O ritmo percussivo ditado por transientes muito curtas dificulta a remoção de ruído. Tanto TFBT quanto SS deixam o som levemente mais "abafado". Contudo, devido à capacidade adaptativa da TFBT, o áudio soa menos "sujo" que o tratado com SS. Semelhante à amostra anterior, se não fossem pelos artefatos inseridos, WT também teria um resultado bom, uma vez que não sofre dos problemas observados no uso da TFBT e SS, mencionados aqui.
- [9 – 28] As notas harmônicas que aparecem acompanhadas da percussão (que continua mais ou menos da mesma forma) são um pouco prejudicadas tanto por TFBT quanto por SS. Soam artificialmente "metalizadas" (talvez pela perda de algumas de suas componentes frequenciais). O desempenho da WT é bom, havendo pouco ruído residual e artefatos.
- [28 – final] As notas de baixa frequência exclusivas deste subtrecho não são perceptivelmente afetadas. WT mantém o bom desempenho.

Linkin Park – What I've Done

- [0 – 3] TFBT é o de melhor desempenho nesse subtrecho, onde áudio original e ruído se confundem. SS remove o ruído, mas exibe o tradicional ruído musical. E WT é insatisfatório, mantendo uma boa quantidade de ruído em todas as frequências.
- [3 – 7] TFBT mantém a clareza das notas do piano neste subtrecho, removendo todo o ruído. SS tem desempenho razoável, em virtude da presença do ruído musical. WT é péssimo neste subtrecho, mantendo muito ruído e ainda acrescentando outros mais.
- [7 – 16] Tanto TFBT quando SS demonstram desempenho insatisfatório neste subtrecho. Ambas falham em manter limpo o som do prato da bateria que dita o ritmo. Era de se esperar o mau desempenho, em virtude de ser uma estrutura muito difícil de ser recuperada[13]. Esses problemas não são observados com o uso do WT, mas seu mau desempenho na remoção do ruído acaba por não compensar também.
- [16 – 22] Neste subtrecho, o ritmo é mais fortemente marcado pela bateria. E todas as técnicas continuam com desempenho ruim, pelos mesmos motivos já citados.

[13] Comportamento parecido é notado em *codecs* de áudio do tipo *lossy*, como o MP3, quando configurados no modo de baixa qualidade (baixo *bitrate*).

- [22 – final] Subtrecho com espectro muito cheio, e com muitos sons naturalmente distorcidos, é difícil avaliar as técnicas. O espectro do sinal tratado com TFBT é ainda assim mais limpo.

OneRepublic – Apologize

- [0 – 8] TFBT e SS cumprem o papel de remover o ruído, porém ambas deixam o som "lavado". Isto se dá devido ao corte das frequências mais altas do sinal, uma vez que possuíam uma magnitude menor (próxima a do ruído). SS adicionalmente deixa alguns artefatos perceptíveis neste trecho mais calmo. WT só consegue ter sucesso na remoção do ruído de frequência mais alta.
- [8 – 16] TFBT e SS tem desempenho muito ruim neste subtrecho. Além do ataque mais forte da percussão ter sido sensivelmente alterado, o som das castanholas é completamente descaracterizado. Nenhuma das duas técnicas consegue remover o ruído em meio as transientes tão rápidas das castanholas. Em ambas o áudio soa igualmente "mascado". SS ainda fica um pouco atrás (se comparado a TFBT) pelo ruído musical presente. A menos dos artefatos típicos da WT, o seu desempenho poderia ser considerado o melhor dentre as técnicas, não sofrendo de nenhum dos problemas citados.
- [16 – final] TFBT e SS tem dificuldade em manter o som do violino que aparece nesse trecho, com notas harmônicas mais longas. WT mantém o mesmo desempenho.

The Verve - Bitter Sweet Symphony

- [0 – 12] Subtrecho com características muito próximas da faixa "*Square One*" anteriormente comentada. Portanto, as técnicas se comportam de maneira similar. TFBT e SS com desempenho parecido, removendo bem o ruído. Ao contrário disso, WT é ineficaz nesta tarefa.
- [12 – 46] TFBT e SS descaracterizam bastante o violino que faz as notas que tornaram essa musica internacionalmente conhecida. Os harmônicos de alta frequência desaparecem, contribuindo para dar uma aparecia mais "abafada" ao som. WT mantém o péssimo desempenho.
- [46 – final] TFBT e SS tem melhor desempenho neste subtrecho, em virtude do espectro mais preenchido. Apesar de o som perder o "brilho" com ambas as técnicas, o

resultado não é dos piores. A WT é reservada o mesmo comentário já dito anteriormente sobre ela.

The Verve - Sonnet

- [0 – 10] Excelente amostra de teste. O belo som dos acordes bem marcados pelo violão de cordas de aço presente na amostra original é destruído por todas as técnicas. Cada uma com seus problemas. Porém todas reconhecidamente ruins. Dentre todas as amostras, talvez a que mais possua ruído musical, provado pela SS. Os artefatos inseridos pela WT soam como "gotas de chuva caindo sobre uma poça de água".
- [10 – final] TFBT e SS melhoram seu desempenho neste subtrecho em função do áudio ser menos "sutil" (mais encorpado). Ambas mantêm as notas de baixa frequência. TFBT, porém, deixa o áudio mais limpo que SS, que "masca" um pouco nas transientes. WT prejudica muito pouco as transientes, mas deixa muito ruído e insere artefatos em demasia no sinal.

The Verve - The Drugs Don't Work

- [trecho único] Definitivamente o *"killer sample"*[14] dentre todas as amostras. O motivo disso, já citado na sua análise da sessão anterior, é o complexo som do violão de cordas de aço em conjunto com as notas contínuas (com timbre de violino, mas provavelmente produzidas por um teclado). Todo o brilho das notas provenientes dos violões é completamente destruído por todas as técnicas. Apenas as notas que fazem a base são relativamente preservadas pela TFBT e SS. WT tem provavelmente seu pior desempenho dentre todas as amostras testadas, preservando apenas as baixas frequências e ainda assim adicionando muitos artefatos.

Vivaldi - Flute Concerto

- [trecho único] TFBT e SS preservam relativamente bem o som do instrumento solista (a flauta), mas removem parte do brilho da base feita pelo cravo (a parte "metálica" das notas é perdida), em virtude de serem representadas por estruturas tempo-frequência de magnitude mais baixa. WT é bastante ineficiente, deixando ruídos bastante perceptíveis ao longo de quase todo o espectro.

[14] Em uma tradução livre, uma "amostra assassina". Comum em testes de *codecs* de áudio, é uma amostra que possui características que evidenciam fortemente os pontos negativos do algoritmo que processa o audio, seja ele um *codec* (como o MP3, AAC, etc.) ou um algoritmo de remoção de ruídos, objeto dessa pesquisa.

4.2.5 Avaliações

Nesta sessão, procederemos a uma série de três análises: objetiva, subjetiva e computacional. No primeiro caso, serão usadas métricas de medida de qualidade/distorção clássicas na área de processamento de sinais. Como será mostrado, o SNR favorece determinadas técnicas, mas os testes subjetivos, com ouvintes classificando e comparando o que estão ouvindo, às vezes dão resultados diferentes. Por fim, a complexidade computacional das técnicas difere muito, e, apesar de não ser o foco dessa dissertação, alguns aspectos serão abordados.

4.2.5.1 Objetivas

Quando se tratam de avaliações ditas objetivas, de sinais de áudio, a medida mais usada, senão a única, é o SNR. Nas mais diversas publicações na área, a análise objetiva da qualidade é feita através do SNR [3][4][7][17].

É importante destacar, no entanto, que um bom SNR não é uma indicação irrefutável de qualidade do sinal. Apesar de muito usado, é preciso cautela uma vez que não raro são descritos casos na literatura em que a percepção subjetiva da qualidade contradiz o SNR [3]. Em outras palavras, um sinal tratado com uma técnica com SNR melhor que outra, pode ser considerado de pior qualidade por um teste subjetivo (na avaliação auditiva de um ser humano).

Tabela 4.9 SNR antes e depois da aplicação de cada técnica sobre cada amostra

Nº	Amostra	SNR Antes (dB)	SNR Depois (dB)		
			Block Thresholding	***Spectral Subtraction***	***Wavelet Thresholding***
01	*Coldplay - Life In Technicolor II*	26,40	24,43	9,12	26,68
02	*Coldplay - Square One*	20,73	26,88	9,12	25,62
03	*Keane - Somewhere Only We Know*	29,63	26,16	9,12	27,56
04	*Keane - Untitled 1*	27,71	25,08	9,11	26,48
05	*Linkin Park - What I've Done*	26,62	26,21	9,11	26,28
06	*OneRepublic - Apologize*	27,80	29,45	9,12	31,48
07	*The Verve - Bitter Sweet Symphony*	22,84	24,86	9,12	24,39
08	*The Verve - Sonnet*	18,44	22,51	9,02	21,39
09	*The Verve - The Drugs Don't Work*	9,02	17,81	8,68	14,42
10	*Vivaldi - Flute Concerto*	26,90	26,82	9,12	27,49

A Tabela 4.9 sumariza o cálculo do SNR de cada amostra tratada por cada uma das técnicas. Pelos motivos já citados na sessão deste capítulo que lida com sinais sintéticos, o cálculo do SNR para a Spectral Subtraction é reconhecidamente “falho” uma vez que o sinal filtrado sofre ganho em relação ao sinal original, o que leva a sensíveis alterações no cálculo.

Assim sendo, a referida técnica não contará com a etapa de avaliação objetiva. Este fato não ocorre com as demais técnicas.

Uma análise cuidadosa da Tabela 4.9 revela algumas informações interessantes. Há casos, como o das amostras 03, 04 e 05, que nenhuma das técnicas conseguiu melhorar o SNR da amostra ruidosa. Há também casos como o da amostra 09, que a Block Thresholding e a Wavelet Thresholding conseguiram melhorar significativamente o SNR.

4.2.5.2 Subjetivas

Uma etapa importante do processo de avaliação de uma técnica de remoção de ruídos em áudio é submeter amostras processadas pelo algoritmo à apreciação de ouvintes. Isso permite que se estabeleça uma "opinião média" sobre o eficiência/qualidade de cada abordagem. Antes de exibir e comentar os resultados, será descrito o procedimento usado na elaboração dos testes.

4.2.5.2.1 Do Procedimento de Testes

Foi solicitado a cada ouvinte que escutasse com atenção a 5 "versões" de cada uma das amostras, a saber: a original, a ruidosa, as tratadas com *Block Thresholding*, *Spectral Subtraction* e por fim com a *Wavelet Thresholding*, nesta sequência.

Todos os testes foram conduzidos (pelo autor), a fim de evitar possíveis falhas. O ouvinte poderia ouvir quantas vezes quisesse cada uma das versões, porém não fora mencionado para ele quaisquer informações sobre as técnicas usadas no processo. Apesar da possibilidade deste ouvir quantas vezes quisesse cada uma das versões, na prática, normalmente, isso não foi necessário. Cada avaliação levou em média 30 minutos para ser executada.

À medida que o ouvinte ia escutando cada faixa, ele ia avaliando cada técnica segundo três aspectos diferentes:

- **Qualidade Geral**: Dos três quesitos, provavelmente o mais subjetivo. O ouvinte atribuiu uma nota que poderia ir de 1 (ruim/menor qualidade) a 5 (ótimo/maior qualidade). O conceito de "qualidade" neste contexto está diretamente ligado ao quanto o ouvinte apreciou o áudio tratado pela técnica em questão, numa observação geral.
- **Perda de Detalhes**: Uma vez que todas as técnicas inevitavelmente provocam alguma perda de detalhes no áudio original, foi pedido que os ouvintes tentassem quantificar essa perda, numa escala de 1 (poucas perdas/mais próximo do original) a 5 (muitas

perdas/mais distante do original). Essa comparação foi possível uma vez que lhes foi dada a possibilidade de ouvir o amostra original (sem adição de ruído).

- **Ruído Residual e/ou Artefatos**: Outro fato a respeito das diferentes abordagens é que muitas delas acabam por não remover 100% do ruído e/ou introduzem distorções no áudio (uma espécie de "efeito colateral"), aos quais foram denominados ruído residual e artefatos, respectivamente. Assim, foi solicitado que tentassem quantificar, numa escala de 1 (pouco ruído residual/artefatos) a 5 (muito ruído residual/artefatos).

A Figura 4.21 reproduz a ficha com a tabela de avaliação subjetiva que foi entregue a cada ouvinte no processo de avaliação. O campo "Nome" foi usado meramente por uma questão de controle, não tendo qualquer influência na avaliação.

Ficha de Avaliação Subjetiva

Qualidade Geral	1 (Péssimo) - 2 (Ruim) - 3 (Razoável) - 4 (Bom) - 5 (Ótimo)
Perda de Detalhes	1 (Pouco) a 5 (Muito)
Ruído Residual e/ou Artefactos	1 (Pouco) a 5 (Muito)

Nome:

Nº	Amostra	Qualidade Geral			Perda de Detalhes			Ruído Residual e/ou Artefactos		
		BT	SS	WT	BT	SS	WT	BT	SS	WT
1	Coldplay - Life In Technicolor II									
2	Coldplay - Square One									
3	Keane - Somewhere Only We Know									
4	Keane - Untitled 1									
5	Linkin Park - What I've Done									
6	OneRepublic - Apologize									
7	The Verve - Bitter Sweet Symphony									
8	The Verve - Sonnet									
9	The Verve - The Drugs Don't Work									
10	Vivaldi - Flute Concerto									

Figura 4.21 Ficha de Avaliação Subjetiva, entregue a cada ouvinte

No que diz respeito aos ouvintes envolvidos, todos eram jovens adultos e nenhum deles possuía conhecimento técnico em análise de áudio. Nenhum dos membros diretamente envolvido neste trabalho participou dos testes, uma vez que isso poderia influenciar (inadequadamente) os resultados.

Quanto ao hardware envolvido, em todos os testes foi usado o mesmo fone de ouvido, de qualidade profissional (AKG, modelo K-44), possuindo uma boa isolação de quaisquer ruídos externos (do ambiente). A placa de som variou, porém considerou-se que isto seria

irrelevante ao processo, uma vez que os conversores A/D empregados são de pelo menos 16 bits por canal. Quanto ao *software* que reproduziu o áudio (o *player*) foi, em todos os casos, o foobar2000, sem qualquer tipo de pós-processamento adicional (como equalização).

4.2.5.2.1 Resultados e Comentários

Uma vez colhidos os resultados das 10 pessoas que participaram como voluntários nos testes, pode-se sumarizar os resultados da média e variância, as quais encontram-se na Tabela 4.10 e Tabela 4.11, respectivamente.

Tabela 4.10 Valores médios das avaliações subjetivas

Média										
Nº	Amostra	Qualidade Geral			Perda de Detalhes			Ruído Residual e/ou Artefactos		
		BT	SS	WT	BT	SS	WT	BT	SS	WT
1	Coldplay - Life In Technicolor II	4,22	3,11	1,89	1,78	2,11	2,78	1,11	2,33	3,89
2	Coldplay - Square One	4,22	3,33	2,11	1,56	1,89	2,89	1,56	2,44	3,44
3	Keane - Somewhere Only We Know	4,22	3,78	2,89	1,44	1,44	2,00	1,00	1,89	3,33
4	Keane - Untitled 1	4,11	3,33	2,44	1,56	2,00	2,67	1,56	2,67	3,33
5	Linkin Park - What I've Done	3,78	3,11	2,11	1,89	2,00	2,44	2,00	2,67	3,89
6	OneRepublic - Apologize	4,00	2,67	2,11	1,78	2,00	2,33	1,56	3,11	4,00
7	The Verve - Bitter Sweet Symphony	3,33	3,22	1,56	2,44	2,00	2,78	1,78	2,56	4,44
8	The Verve - Sonnet	3,44	2,78	1,44	2,22	1,67	3,56	1,56	3,00	4,78
9	The Verve - The Drugs Don't Work	3,22	2,89	1,22	2,67	2,33	4,22	1,56	3,56	4,78
10	Vivaldi - Flute Concerto	4,33	3,33	1,78	1,22	1,78	2,78	1,33	2,44	4,11

Tabela 4.11 Variância das notas atribuídas a cada amostra nas avaliações subjetivas

Variância										
Nº	Amostra	Qualidade Geral			Perda de Detalhes			Ruído Residual e/ou Artefactos		
		BT	SS	WT	BT	SS	WT	BT	SS	WT
1	Coldplay - Life In Technicolor II	0,44	0,86	0,11	0,69	1,36	0,94	0,11	1,00	0,36
2	Coldplay - Square One	0,69	0,75	0,36	0,78	0,61	0,36	0,53	1,03	0,53
3	Keane - Somewhere Only We Know	1,19	0,44	0,61	0,53	0,53	0,75	0,00	0,36	0,75
4	Keane - Untitled 1	0,86	0,50	0,53	0,28	1,00	1,25	0,53	0,50	1,00
5	Linkin Park - What I've Done	0,94	0,61	0,86	0,61	0,25	0,78	1,50	0,75	0,86
6	OneRepublic - Apologize	0,50	0,50	0,61	0,69	0,50	0,50	0,53	0,36	1,00
7	The Verve - Bitter Sweet Symphony	0,75	0,94	0,53	1,28	1,25	0,94	1,19	1,03	0,28
8	The Verve - Sonnet	1,78	0,44	0,53	0,44	0,50	0,78	1,03	0,50	0,44
9	The Verve - The Drugs Don't Work	0,69	0,61	0,19	1,25	0,50	0,44	0,28	1,03	0,19
10	Vivaldi - Flute Concerto	0,50	1,25	0,69	0,19	1,19	0,19	0,50	1,03	0,61

As cores verde, amarelo e vermelho, destacam, em ordem decrescente os resultados, ou seja, do melhor para o pior. Foi unânime a atribuição dos melhores resultados no quesito "Qualidade Geral" a *Block Thresholding*. Em segundo lugar, a *Spectral Subtraction*. Em último, a *Wavelet Thresholding*.

Uma observação interessante é que para algumas amostras, como a 05, 07 e 09, a qualidade apontada pelos testes indica uma proximidade grande entre as técnicas *Block Thresholding* e *Spectral Subtraction*.

De maneira geral, a *Wavelet Thresholding* não apenas se mantém em último lugar, como também evidencia uma "distância" maior do segundo lugar do que a distância que o segundo mantém do primeiro.

No que diz respeito à "Perda de Detalhes", segundo os ouvintes, a *Spectral Subtraction* consegue perder menos detalhes que a *Block Thresholding* em algumas amostras, a saber, 07, 08 e 09. Essas três amostras têm em comum a presença de sons muito sutis e difíceis de serem mantidos pelos algoritmos de remoção de ruído, conforme comentado na sessão 4.2.4.3. Não é coincidência, portanto, que tais amostras tenham conseguido também as menores notas no quesito "Qualidade Geral".

Quanto à avaliação do "Ruído Residual e/ou Artefatos", novamente a *Block Thresholding* foi a melhor, seguida da *Spectral Subtraction* e *Wavelet Thresholding*. De fato, os ouvintes concordam que a *Block Thresholding* tem uma incidência mínima de ruído residual e quaisquer artefatos, especialmente o ruído musical, conforme comentado nas explicações teóricas da técnica. *Block Thresholding* teve notas medianas enquanto a *Wavelet Thresholding* evidenciou uma forte presença de ruídos residuais e artefatos, especialmente nas últimas amostras, onde de fato é possível perceber mais isso.

A Tabela 4.11, com as variâncias, tem como propósito evidenciar as dispersões das notas em torno da média. Pode-se dizer que, quanto maior a variância, mais os ouvintes discordaram entre si em determinado quesito/amostra/técnica. Por exemplo, no quesito "Qualidade Geral" da amostra 08, da técnica *Block Thresholding*, as notas variaram mais que as demais. De fato, uma amostra difícil de ser analisada. Outro valor interessante é a variância das notas atribuídas à amostra 05, tratada com a *Block Thresholding* e avaliada no quesito "Ruído Residual e/ou Artefatos". Houve um valor relativamente alto, mas coerente, uma vez que essa amostra possui muito ruído que naturalmente faz parte da música, o que pode ter confundido o ouvinte.

4.2.5.3 Computacionais

Uma análise computacional completa dos algoritmos de remoção de ruído em áudio por si só já dariam uma inteira dissertação de mestrado [25]. Mesmo não sendo o foco deste trabalho, serão comentados alguns aspectos observados ao longo da pesquisa. Estão divididos em duas categorias: desempenho computacional e outros comentários.

4.2.5.3.1 Desempenho Computacional

Quando se fala em desempenho computacional, a principal medida de esforço real de processamento é o tempo. A Tabela 4.12 exibe informações sobre cada amostra, bem como os tempos de processamento de cada técnica. No que diz respeito ao *hardware* de teste, os algoritmos foram executados num PC com 2 GB de RAM, processador Intel Pentium® Dual Core 2.4 GHz. Quanto ao *software*, os algoritmos foram escritos em MATLAB® M-Code, rodando na versão 2009b, sobre o Windows XP SP3.

Tabela 4.12 Estatísticas de tempo de processamento das amostras

Amostra	Duração (em min.)	*Samples*	*Block Thresholding*		*Spectral Subtraction*		*Wavelet Thresholding*	
			Tempo Proc. (seg)	Veloc.	Tempo Proc. (seg)	Veloc.	Tempo Proc. (seg)	Veloc.
Coldplay - Life In Technicolor II	1:12.148	3.181.746	98,452	0,63	5,039	12,33	6,613	9,40
Coldplay - Square One	0:16.230	715.740	22,418	0,72	0,885	18,34	1,040	15,61
Keane - Somewhere Only We Know	0:23.021	1.015.234	31,020	0,74	1,227	18,76	1,468	15,68
Keane - Untitled 1	0:45.699	2.015.333	62,968	0,73	2,602	17,56	2,941	15,54
Linkin Park - What I've Done	0:40.310	1.777.677	54,900	0,73	2,245	17,96	2,534	15,91
OneRepublic - Apologize	0:32.417	1.429.582	45,701	0,71	1,782	18,19	2,068	15,68
The Verve - Bitter Sweet Symphony	1:06.981	2.953.848	93,756	0,71	4,098	16,34	4,197	15,96
The Verve - Sonnet	0:21.887	965.225	30,071	0,73	1,185	18,47	1,385	15,80
The Verve - The Drugs Don't Work	0:24.079	1.061.867	33,429	0,72	1,353	17,80	1,520	15,84
Vivaldi - Flute Concerto	0:23.789	1.049.091	32,635	0,73	1,262	18,85	1,509	15,76

Analisando a tabela, além do tempo de processamento, há uma medida de "velocidade" que é dada pela razão entre a duração da amostra e o tempo de processamento. Este valor mede quão rápido é o algoritmo se comparado ao tempo real de execução da amostra. Em outras palavras, valores acima de 1 indicam que, se levado em consideração apenas o tempo de execução (uma vez que há diversos outros fatores envolvidos), o algoritmo é rápido o suficiente para operar em tempo real, sob o referido hardware de teste.

Há uma discrepância na medida do tempo que pode ser observada na primeira amostra, sendo os algoritmos consideravelmente mais rápidos que nas demais amostras. O motivo disto é desconhecido e pode envolver inúmeros fatores internos do MATLAB.

Contudo, se desconsiderada as estatísticas dessa amostra, as demais possuem coerência com a média.

De imediato pode-se observar a considerável complexidade da *Block Thresholding* frente às outras duas técnicas. Da maneira em que ela se encontra implementada, seu uso em tempo real é inviável. Quanto a *Spectral Subtraction*, pode-se dizer que ela é, na média, um pouco mais rápida que a *Wavelet Thresholding*. Contudo, se considerados apenas o tempo de processamento, ambas são rápidas o suficiente para processamento em tempo real.

4.2.5.3.2 Outros Comentários

No que diz respeito à já comentada aplicação em tempo real, outros fatores estão envolvidos na sua viabilidade. Nenhuma das técnicas opera sobre cada amostra pontual do sinal, mas sim sobre um conjunto delas, ou um *frame* (janela), como é frequentemente chamado. Isso implica na obrigatoriedade de um *buffer* e na introdução de um *delay* no sistema. O tamanho de cada *frame* depende de cada técnica e de ajustes empíricos (ou seja, ajustes que dão melhores resultados na prática). A *Block Thresholding* seleciona automaticamente o tamanho do *frame*, enquanto na *Spectral Subtraction* este é um parâmetro fixo configurável. A *Wavelet Thresholding*, tal qual se encontra implementada, opera sobre o sinal como um todo, não efetuando o processamento apenas sobre parte do sinal, o que impossibilita o uso direto da técnica em tempo real. Mas, modificações no algoritmo poderiam ser feitas a fim de possibilitar isso [25].

Capítulo 5: Conclusão

5.1 Conclusões Gerais a Respeito das Técnicas

Após um extenso capítulo de testes, é possível tirar algumas conclusões a repeito de cada técnica de remoção de ruídos. Nesta sessão, cada abordagem será comentada, com respectivos pontos positivos e negativos que puderam ser verificados.

5.1.1 *Spectral Subtraction*

Foi, de certa maneira, surpreendente ver uma técnica que data dos anos 70 [17], após a modificação proposta, adaptando-a para sinal áudio musical, ter obtido resultados que podem ser considerados satisfatórios. *Spectral Subtraction* parte de um princípio relativamente simples, e, se considerados os requisitos da técnica (especialmente a presença de um trecho inicial com amostras pontuais suficientes contendo apenas ruído), o desempenho é bom.

Pode-se perceber que, apesar do piso espectral evitar que os coeficientes frequenciais adquiram valores negativos, isso não é suficiente para eliminar o ruído musical, que ainda pode ser percebido, especialmente em algumas amostras. Um fato interessante é que, na opinião geral dos ouvintes nos testes subjetivos, esse fato foi pouco mencionado. Ao que parece, o ruído musical (que se encontra num nível relativamente baixo nas amostras testadas, porém perceptível) não incomodou muito a apreciação do áudio pelos ouvintes. Ao menos não tanto quanto as baixas avaliações atribuídas ao quesito “Ruído Residual”, bastante presente nas amostras processadas com a *Wavelet Thresholding*.

Um aspecto muito positivo dessa técnica é que, em virtude da simplicidade, tem um algoritmo bastante rápido, conforme mostrado na Tabela 4.12. Isso se dá porque a parte em que poderia ser consumido mais processamento – as transformações diretas e inversas no domínio da frequência – possui algoritmo rápido (FFT – *Fast Fourier Transform*).

A técnica, tal qual se encontra, mantém a mesma estimativa do ruído ao longo do processamento de toda a amostra. Isso não veio a ser um problema uma vez que o ruído foi assumido estacionário. No entanto, caso as estatísticas do ruído mudassem ao longo da amostra, sem dúvidas a técnica iria falhar, dado que não é adaptativa.

No algoritmo da *Spectral Subtraction*, dois parâmetros precisam ser ajustados de maneira conjunta para que a técnica funcione bem, a saber, α e β, comentados na sessão 3.1.1. Observou-se que α e β precisam ser sintonizados para outros valores de potência de ruído presente na música no intuito de se obter um resultado próximo ao ponto ótimo da

técnica. Pensando num contexto prático, uma interface poderia disponibilizar para o usuário o controle desses dois parâmetros e este, por experimentação, não teria dificuldade em encontrar os valores adequados.

5.1.2 *Time-Frequency Block Thresholding*

Numa opinião geral, sem dúvidas a técnica com o melhor desempenho dentre as testadas. O objetivo de eliminar o ruído musical foi, de fato, alcançado. A abordagem usada é muito eficiente na remoção do ruído, enquanto tenta manter as estruturas do áudio original.

Observou-se que, na tentativa de remover o máximo possível de ruído, a técnica acaba por deixar o áudio "excessivamente limpo", ou seja, as estruturas mais sutis, que dão "brilho" ao som (como os harmônicos de baixa magnitude e alta frequência de alguns instrumentos) desaparecem. O autor chama este cenário de "remoção total do ruído" [7]. Mas é possível que, num cenário de remoção parcial (percentual, porém, indefinido) do ruído, pode-se equilibrar melhor a quantidade de ruído removida e as perdas de detalhes sofridas.

Conforme mostrado da sessão 3.2, a técnica é robusta por ser adaptativa. Contudo, ela necessita do valor do desvio padrão (ou uma estimativa deste) do ruído. Em testes preliminares, pode-se perceber que a quantidade de ruído removida tem relação direta com esse parâmetro, o qual poderia, inclusive, agir como um "controle" da quantidade de ruído removida[15]. O valor desse parâmetro poderia ser estimado e passado ao algoritmo ou, num contexto prático, encontrado experimentalmente. O usuário poderia ir testando valores até que encontrasse um valor que promovesse uma remoção adequada do ruído, estando assim próximo do valor real do desvio padrão do ruído.

Contudo, há um problema com o procedimento comentado no parágrafo anterior. Como mostra a Tabela 4.12, *Block Thresholding*, é uma técnica com custo computacional elevado, atribuído principalmente à busca do melhor ladrilhamento do espaço tempo-frequência. Esta busca é exaustiva, apesar de trabalhar com um conjunto limitado de possibilidades. No algoritmo de teste, há 15 possibilidades diferentes de blocos para cada macrobloco. Esse alto custo computacional é limitante para a implementação de um ambiente interativo, onde o usuário poderia, por tentativa e erro, encontrar um valor apropriado. É digno de nota, porém, que o algoritmo fornecido pelo autor é uma prova conceitual, e, portanto, não computacionalmente ótima. É razoável pensar que há diversas possibilidades de otimização neste campo.

[15] Ou seja, suponha que o valor do desvio padrão do ruído seja X. Se invés de X fosse fornecido o valor 0.5X, pode-se esperar que seja possível ouvir uma quantidade significativa de ruído residual. O algoritmo "guiou-se" pela estimativa dada na hora de calcular os coeficientes de atenuação.

5.1.3 *Wavelet Thresholding*

A proposta de testar o desempenho da técnica *Wavelet Thresholding* na remoção de ruído em sinal de áudio musical foi motivada pelo seu reconhecidamente bom desempenho em testes com sinais sintéticos apresentados na literatura. No entanto, os resultados foram muito aquém do esperado. O principal problema encontrado foi a forte presença de artefatos no sinal tratado. Essas distorções, em alguns casos, incomodaram muito mais que o próprio ruído que a técnica se propunha a remover.

No que diz respeito a aspectos computacionais, no entanto, pode-se apresentar o bom desempenho computacional. O algoritmo é bastante rápido se comparado a *Block Thresholding*.

5.2 Projetos Futuros

A abordagem adotada na técnica *Time-Frequency Block Thresholding* foi elaborada de tal maneira que fosse robusta e não particularizada para casos específicos. No artigo [7], o autor faz, por exemplo, testes com amostras de voz, bem como de áudio instrumental (apesar de serem amostras bastante curtas). É provável que a técnica pudesse ser particularizada para alguns casos, gerando resultados ainda melhores. Por exemplo, se fosse adicionado um modelo psicoacústico que ajudasse a decidir a intensidade a partir da qual um ruído poderia ser perceptível, e se trabalhasse sobre esse limiar, talvez os resultados pudessem ser melhores, especialmente no que se refere a manter as estruturas mais frágeis (de menor magnitude) do sinal.

Outro fato ainda a respeito da *Time-Frequency Block Thresholding* é que, apesar de ter sido aplicada usando a Transformada de Fourier, toda a teoria envolvida é mais geral, podendo, inclusive, ser aplicada sobre coeficientes *wavelets*. Os efeitos dessa mudança de domínio poderiam ser melhor investigados.

No que diz respeito a *Spectral Subtraction*, poderia ser desenvolvida uma técnica semelhante à empregada pelo algoritmo VAD a fim de detectar regiões de relativo silêncio na música, e, a partir desses trechos, atualizar as estimativas do ruído (seu espectro).

Existem várias outras abordagens de remoção de ruídos que não foram testadas neste trabalho. Algumas, típicas da área de processamento de imagens, poderiam dar resultados satisfatórios se adaptadas para sinais unidimensionais, em particular áudio musical. Uma dessas técnicas é a que visa suavizar imagens, porém preservando as bordas, como a Difusão Anisotrópica. A ideia é preservar as bordas que, no caso de áudio, representariam os ataques. Entretanto, cabe ressaltar que uma transição rápida numa imagem, caracterizando uma borda,

não necessariamente terá seu correspondente em um sinal de áudio. Portanto, a técnica precisa ser avaliada e adaptada convenientemente para tratar sinais 1D reais e não sinais 1D sintéticos, onde é possível haver transições rápidas, em um período de amostragem.

Referências Bibliográficas

[1] D. Salomon, “Data Compression – The Complete Reference”, 4th Edition, Springer, 2007.

[2] D. L. Donoho, I. M. Johnstone, “Adapting to Unknown Smoothness via Wavelet Shrinkage”, Journal of the American Statistical Association, Vol. 90, 1995.

[3] M. Welk, A. Bergmeister and J. Weickert, “Denoising of Audio Data by Nonlinear Diffusion”, Lecture Notes in Computer Science, Vol. 3459, Springer, Berlin, 598–609, 2005.

[4] S. V. Vaseghi, “Advanced Digital Signal Processing and Noise Reduction”, 3rd Edition, Wiley, 2006.

[5] A. Papoulis, S. Unnikrishna Pillai, “Probability, Random Variables and Stochastic Processes”, McGraw Hill Higher Education, 4th Edition, 2002.

[6] B. Widrow, S. Stearns, “Adaptive Signal Processing”, Prentice Hall, 1985.

[7] G. Yu, S. Mallat, “Audio Denoising by Time-Frequency Block Thresholding”, IEEE Transactions On Signal Processing, Vol. 56, No. 5, May 2008.

[8] G. Yu, Personal Home Page. Disponível em <http://www.cmap.polytechnique.fr/~yu/research/ABT/samples.html> (Acesso 01/11/2009).

[9] H. Suzuki, J. Igarashi, and Y. Ishii, “Extraction of Speech in Noise by Digital Filtering”, J. Acoust. Soc. of Japan, Vol. 33, No. 8, pp. 105-411, Aug. 1977.

[10] S. Boll, “Suppression of Noise in Speech Using the SABER Method”, ICASSP, pp. 606-609, April 1978.

[11] S. Boll, “Suppression of Acoustic Noise in Speech Using Spectral Subtraction”, IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. 27, Issue 2, pp. 113-120, April 1979.

[12] R. A. Curtis, R. J. Niederjohn, “An Investigation of Several Frequency-Domain Methods for Enhancing the Intelligibility of Speech in Wideband Random Noise”, ICASSP, pp. 602-605, April 1978.

[13] E. Zavarehei, Disponível em <http://dea.brunel.ac.uk/cmsp/Home_Esfandiar> (Acesso 13/05/2009).

[14] R. M. Udrea, S. Ciochina, "Speech Enhancement Using Spectral Over-subtraction and Residual NoiseReduction", International Symposium on Signals, Circuits and Systems, Vol. 1, pp. 165-168, July 2003.

[15] H. M. Goodarzi, S. Seyedtabaii, "Aplication of Spectral Subtraction for Reducing Industrial Noises", Image and Signal Processing and Analysis, ISPA 2009, pp. 79-83, September 2009.

[16] M. Yektaeian, R. Amirfattahi, "Comparison of Spectral Subtraction Methods Used in Noise Suppression Algorithms", Information, Communications & Signal Processing, pp. 1-4, December 2007.

[17] M. Berouti, R. Schwartz, J. Makhoul, “Enhancement of speech corrupted by acoustic noise”, ICASSP, Vol. 4, pp. 208-211, Apr 1979.

[18] J.O. Smith, “Spectrogram of Speech”, Disponível em <https://ccrma.stanford.edu/~jos/st/Spectrogram_Speech.html> (Acesso 26/11/2009)

[19] C. Stein, “Estimation of the mean of a multivariate normal distribution,” Ann. Statist., vol. 9, pp.1135–1151, 1980.

[20] D. Donoho and I. Johnstone, “Ideal spatial adaptation via wavelet shrinkage,” Biometrika, vol. 81, pp.425–455, 1994.

[21] T. Cai and H. Zhou, “A Data-driven block thresholding approach to wavelet estimation,” Statistics Dept., Univ. of Pennsylvania, Tech. Rep., 2005.

[22] F. Abramovich, T. C. Bailey, T. Sapatinas, “Wavelet Analysis and Its Statistical Applications”, The Statistician, Vol. 49, No. 1, pp. 1-29, 2000.

[23] Mathworks, “Wavelet Toolbox Documentation”, v. 2009b.

[24] G. Strang, “The Search for a Good Basis”, Technical Report, Department of Mathematics, Massachusetts Institute of Technology, 1997.

[25] C. Schremmer, “Wavelets In Real Time Digital Audio Processing: Analysis And Sample Implementations”, Master’s thesis, University of Mannheim, May 2000.

[26] D. L. Donoho, “Denoising by Soft Thresholding”, IEEE Transactions on Information Theory, Vol. 41, pp. 613-627, May 1995.

[27] C. Févotte, B. Torrésani, L. Daudet, and S. J. Godsill, “Sparse Linear Regression with Structured Priors and Application to Denoising of Musical Audio”, IEEE Transactions On Audio, Speech, And Language Processing, Vol. 16, No. 1, January 2008.

[28] Hydrogenaudio.org, “LAME 3.98.2 VBR bitrate test, all -V settings in 0.5 step increments”, Disponível em <http://www.hydrogenaudio.org/forums/index.php?showtopic=67523> (Acesso 31/10/2009).